

Free trade in human sequence data?

British and US institutions have separately challenged the notion that the nucleotide sequences of the human genome should be an open book; have they also signalled the end of the age of innocence for the new biology?

THE project to determine the nucleotide sequence of the human genome is in trouble before it has properly begun. Two separate developments, in Britain and the United States (see *Nature* 354, 96; 14 November 1991), have ensured that. They will also undermine the cozy assumption that people with nucleotide sequence data of any kind will add them, as a matter of routine and in the public interest, to the freely accessible international databanks. And they will further confuse the question of the proper role of basic research in the exploitation of discovery.

The technique from which confusion springs is that of the polymerase chain reaction (PCR), which in principle allows the ad lib replication of single polymeric nucleic acid molecules. Applied to the molecules of messenger RNA in cells from particular tissues, and with the help of reverse transcriptase to turn RNA into DNA, the technique can provide quantities of DNA large enough to be sequenced.

The threat to the human genome project arises because these sequences have been supposed to have commercial value. In the United States, the National Institutes of Health (NIH) have applied for a patent to protect 337 cDNA sequences published earlier this year by Dr Craig Venter and his associates, and derived from brain tissue. Britain's Medical Research Council (MRC) had been planning to follow a different tack, that of charging commercial users fees for access to what it regards as the proprietary information compiled by its Human Genome Mapping Group, whose guru is Dr Sydney Brenner. It is a striking proof of the efficacy of the technique that MRC's modest effort has identified 2,000 genes or gene fragments in only a few months. At that rate, the 100,000 or so genes in the entire genome are plainly within reach.

Customers

Who will pay for access to this data? The most obvious customers are pharmaceutical companies. They might find a set of tissue-specific sequence data useful as pointers to genes whose products, when made abnormal by mutation or otherwise, may be relevant to the treatment of a tissue-specific disease. Alternatively, a comparison between cDNA libraries from healthy and diseased tissue might well point directly to the functional disorder responsible for a disease, and thus to a means of correcting it. Those seeking more effective means of treating schizophrenia, for example, may reasonably calculate that much

is to be learned from cDNA sequences from the *substantia nigra* of sick and healthy people. The potential profits of success are plainly very large. So should not those constructing cDNA libraries, not to mention their paymasters, share in the rewards? That is the question NIH and MRC seem to have answered affirmatively.

Miscalculations

They have miscalculated. Data about cDNA genes and gene fragments churned out by automatic sequence analysers have literally no value in themselves. At the very least, thought is needed to make them meaningful; deciding with which other sequences they are best compared is not, for example, a trivial pursuit. And pharmaceutical companies hoping that sequence data will lead them to new drugs will often have to carry out parallel investigations of their own, perhaps to compare cDNA from diseased tissue with what MRC or NIH have to offer. It is unthinkable, in any case, that any company would embark on the development of an innovative drug, at a potential cost of \$100 million or so, on the basis of cDNA sequences alone.

The much more probable consequence of the fuss about cDNA sequences is that substantial companies will set up their own sequencing projects. With annual research budgets measured in units of \$100 million, an annual payment of £5,000 will seem to many companies a small price to pay for knowing what others know. MRC's belief in the commercial value of its mapping group's data would be more convincing if it had been planning to charge real money for access. Why not £5 million a year? Because even the richest companies would be outraged — and would then be sure to do the work themselves. MRC should also give some thought to what it means by the 'non-commercialization' agreements it would exact from academic researchers: does that mean that academics must refrain from making useful discoveries with MRC's data, or that the intellectual property in all discoveries belongs to MRC and not to them?

NIH's search for patent protection is no more convincing. It is the equivalent in molecular biology of the nineteenth century practice in the United States of letting gold miners stake claims to the mineral resources beneath arbitrarily chosen plots of land. The project to patent Venter's sequences is speculative in just that sense. The sequences themselves are in the public domain, via



GenBank. NIH seems convinced that there is a high chance of its application (under US law a confidential document which NIH has made public) being granted, on the grounds that the gene sequences are novel and have been found by a method which is not obvious. But not all patent offices will agree to offer protection to a cDNA sequence that is nothing more than a description of the natural world. (Newly discovered galaxies are not patented.) And while NIH may have the laudable aim of ensuring that significant sequences are fully exploited, the chances that this first batch of cDNA sequences will offer a route to the treatment of disease cannot be high. The chances of success are probably not very different from those afflicting the gold-rush prospectors.

Both MRC and NIH are thus looking for speculative (but probably insubstantial) benefits from the first flush of data with a new technique. The policies are discreditable because they make a monkey of honest research, and because so little thought has been given to the damage they will do. What, for example, will happen now to the exhortations of those in the academic sector with novel sequence data that they should add them to the public databanks? The case for doing so is strong, for these are eminently databanks in which the whole is greater than the sum of the parts. Sequencers are now advised (by NIH) that they should also seek patent protection and (by MRC) that they should keep the data on their own computers and charge for access. (To be fair, the managers of the public databanks must also now confess that they should have worried more about commercial usage from the outset, and should have worked harder to match the presumed obligation of the academic sector freely to contribute its sequences by a similar obligation upon the world's industrial laboratories.)

But if either NIH or MRC succeeds, the human genome projects will be a more conspicuous casualty. The original ambition was that there would be an international effort to tackle the biggest practical problem to come the way of academic biology. The United States was acknowledged to be the richest source of ideas and also, inevitably, the most generous source of research funds (for work largely but not exclusively carried out in US laboratories), but the contributions of groups elsewhere (notably in France and Japan) were recognized as crucial.

In the absence of an international treaty to regulate the use of the accumulating data, the cheerleaders of international collaboration set up a private organization, HUGO (for Human Genome Organization), to divide the work among would-be participants, informally coordinating national programmes of research, to encourage governments to give the projects backing and to attempt to reach agreement on access to sensitive data — but little except public speaking has been done about that. MRC's plans, by balkanizing the databases, will be the more serious threat to the genome projects. Has anybody weighed the damage that will be done to a brave enterprise against the small change NIH and MRC may, if they are lucky, collect?

It is ironical that these ill-considered steps have been taken by organizations whose remit includes the promotion of the welfare of basic research and which have admirable reputations in that role. Why have they behaved aberrantly? The charitable explanation is that they share the pressure to which individual researchers are increasingly subjected to turn discovery into cash. Mrs Margaret Thatcher, when British prime minister, by perpetually sniping at MRC for having failed to patent Milstein's technique for making monoclonal antibodies, created a climate in which it came to seem a researcher's 'bounden duty' (an authentic phrase) to patent everything in sight. But MRC is a grown-up organization, and should not be terrified by voices from the political grave. NIH's plans are less immediately destructive of the genome projects, but it will be distracting (and costly) if everybody with a new cDNA sequence seeks preemptive patent protection.

What should be done? NIH seems weakly to be hoping that its political masters will hammer out a workable policy (see page 174), but is that really necessary? NIH is better placed than, say, Dr Louis Sullivan to balance the benefits of a patent policy against the damage it will do. Could not NIH, acting independently, decide not to seek patent protection for sequences as such, but only for sequences informed by such a degree of understanding that the package is a plausible basis for some drug treatment? That, in any case, is what the world's patent offices are likely to say. If the patent lawyers, perhaps with a vested interest in the prospect of a delicious growth of business, believe there is doubt, it should not be too difficult to find an obliging congressman to introduce clarifying legislation.

MRC is a more difficult case. Like many other British research organizations, it could use the cash. But there are more civilized ways in which that could be done. Why not turn its present plan inside-out, asking potential customers to contribute financially to the continuing cost of a research programme that will ultimately be of great benefit to them as well as to the research community as a whole? (Three years ago, Professor Walter Gilbert was pushing for just such an arrangement on behalf of the international genome project; his scheme should be looked at again.)

The most urgent need is to revitalize HUGO, and to hammer out the international agreements on access to data, and the residual rights to intellectual property of those who contribute to it, without which the grand conception of the human genome project will be doomed. But that will take time. That is why, in Britain and the United States, the national academies (the Royal Society and the US National Academy of Science) should take it upon themselves to see that MRC and NIH, the first black sheep to present themselves, balance in a more seemly fashion the now conflicting interests that the human genome should be understood and that the potential commercial beneficiaries should pay. Since when has it been a crime that public spending on basic research should prove to be beneficial to the useful arts? □

FDA announces reforms of drug review process

- Agency to use industry contractors for reviews
- Congress raises conflict-of-interest concerns

Washington

OVERWORKED and criticized by activists who say that drug approval delays cost lives, the US Food and Drug Administration (FDA) announced last week major new reforms to speed its reviews of new pharmaceuticals. While the agency will continue to use its own inspectors to review innovative or badly needed drugs, it will hire outside contractors — many of whom are expected to come from the pharmaceutical industry — to review more routine products.

Although the pharmaceutical industry last week hailed the reforms as long-needed changes that will reduce both the cost of new drugs and their delay in reaching market, critics warned that use of industry essentially to police itself risked conflict of interest. And in a 13 November letter to FDA commissioner David Kessler, Representatives John Dingell (Democrat, Michigan) and Henry Waxman (Democrat, California) and Senator Edward Kennedy (Democrat, Massachusetts) raised concerns that the reforms are "thinly veiled efforts to weaken the agency".

Pharmaceutical companies have long complained about the backlog in the FDA drug approval process — on average, it now takes 9.75 years for a drug to be developed and approved. By focusing its internal review efforts on the most important drugs, a new procedure the agency calls "accelerated approval" and aims to have in place by 1994, FDA hopes to reduce that period to 5.5 years. For more routine drugs, which will be reviewed by a combination of internal staff and contractors, FDA intends to cut almost three years off the process of development and approval, to an average of seven years.

In a rare convergence of interests, the new proposals have the effect of answering the demands of both industry and disease activists. AIDS groups, in particular, have been vocal in their criticism of FDA for delays in approving experimental AIDS therapies. But it was the mounting concerns of the biotechnology and pharmaceutical industry that finally goaded the Administration into action. The reforms come not from FDA itself, but from the White House Council on Competitiveness, a body headed by Vice President Dan Quayle that is devoted to ensuring that government does not interfere unduly with industry (see *Nature* 353, 199; 19 September 1991).

But to the reform's Congressional critics, the moves smack of politically moti-

vated deregulation, with potential harm to public health. Despite a series of generic drug scandals that pointed out weaknesses in FDA's internal controls, the agency has remained understaffed and poorly funded in the Administration's annual budget request, the critics note.

Other opponents of the moves argue that FDA will find it difficult, if not impossible, to retain enough conflict-free outside researchers to review the volumes of scientific and clinical-trial data that companies must submit with their drug applications. "There isn't an independent group of contractors out there. They're going to have to go to industry, and I think that's dangerous," says Margaret Mellon, director of the National Wildlife Foundation's Biotechnology Policy Center.

The Competitiveness Council plan states that FDA will "establish safeguards to avoid the appearance of conflict of interest and prevent the opportunity for a company to influence the outcome of the external review." But council officials say that this does not mean new regulations; they believe that existing conflict-of-interest rules are sufficient to ensure that independent reviewers are not materially interested in the drugs they evaluate. They also note that both France and the United Kingdom already rely heavily on outside reviewers in approving new drugs.

The new FDA reforms will also:

- allow Institutional Review Boards — the panels of scientists at universities or companies who currently review research protocols at their own institutions — to approve phase I clinical trials of new drugs under study. This will let FDA safely ignore much of the initial clinical research for new therapies, in favour of focusing on those therapies that have passed the first test and are more likely to become commercial products.

- relax the standards for efficacy to take into consideration the risk to health that delaying a drug for further testing may entail. Based on policies that FDA has already started to use for some AIDS drugs, this change essentially argues that as long as a potentially life-saving therapy is not dangerous, it is better to have it available while researchers determine its efficacy than withhold it for years of further clinical trials.

- take into consideration approval of drugs in other countries. FDA is now negotiating with the European Communities to develop common standards for clinical trial, data formats and animal testing.

By the end of the process, FDA hopes to be able to automatically approve drugs for sale in the United States if they have already been approved by the EC.

Although some of the reforms can be implemented by simply changing internal FDA policy, most will require formal rule-making procedures — including public comment periods — and some will need new congressional legislation before they can take effect. **Christopher Anderson**

SCIENTIFIC MISCONDUCT

Tables turned

Washington

ANYONE who has ever contemplated the sobering prospect of an investigation by Representative John Dingell (Democrat, Michigan) or an inquiry by freelance science cop Walter Stewart can appreciate the irony: Stewart, backed by an attorney representing Dingell's committee, was himself grilled last month in a day-long deposition, and he invoked 'executive privilege' 41 times in refusing to answer questions posed by an attorney representing one of the targets of his investigations.

The case at issue was that of L. Cass Terry, a Medical College of Wisconsin neurologist who is the subject of an investigation by the National Institutes of Health (NIH) misconduct office (see *Nature* 349, 357; 31 January 1991). Terry's lawyer, Matthew Flynn, is pulling out the stops in his client's defence — beyond arguing that Terry was set up by an angry employee, Flynn is charging that NIH's misconduct policies are illegal, and that Stewart violated his client's rights by leaking secret information about the case to the press.

In the course of Flynn's questioning, which veered wildly through Stewart's undistinguished history as an NIH physicist to his motivations for switching to misconduct as a research subject, Stewart managed to forget the name of the NIH institute that employs him (it is the National Institute of Diabetes and Digestive and Kidney Diseases), and was spared only by an objection from his lawyer from having to discuss whether he suggested that a certain whistle-blower hide a microphone in her underwear. But throughout, he defended his misconduct work as a simple research pursuit, as valid as any other.

Charles Tiefer, Stewart's congressional attorney, says that executive privilege (also known as 'speech and debate privilege') is designed to protect legislators from being sued for possibly slanderous things they may say in session. The Supreme Court, he says, has ruled that the privilege can be extended to the working of any congressional investigation, such as the Dingell probes in which Stewart participated. Without the opportunity to occasionally hide behind this legal protection, he says, Congress would be unable to conduct investigations. **Christopher Anderson**

More questions than answers

Bethesda, Maryland

SEEKING to interject a little reason and experience into what has become a heated international fracas over the patenting of human genes, the National Institutes of Health (NIH) last week asked a dozen patent and technology-transfer experts to argue the issues in public. Fourteen hours of debate later, the experts appeared to have vindicated NIH for launching the controversy, even if they were no closer to agreeing whether it is desirable, or even possible, to patent the genetic sequences under question.

If there were still doubt that the question of whether unidentified gene fragments should be patentable has become the most contentious scientific debate of the day, the NIH meeting should have banished it. A full house of observers sat through a day and half a night of esoteric argument on the subtleties of international patent law until NIH had to shut the building for the night. But many left wondering just what to conclude. While scientists had argued that the debate is essentially legal, and lawyers had explained that it is really all up to the US patent office, one could not help noticing that the audience included several patent office officials asking questions and taking careful notes.

Since news emerged last month that NIH had quietly filed a patent application for some 350 cDNA fragments of unknown function and physical significance, the agency has been accused of breaking the international genome project's code of free data exchange, as well as of undermining US biotechnology industry by attempting to put frivolous licensing hurdles

in the path of commercialization (see *Nature* 353, 485; 10 October, 1991). Although most of the industry representatives and lawyers who spoke out at the meeting thought that patents on such gene fragments would do industry more harm than good, they also noted that much the same had been said of the first attempts to patent recombinant DNA technology in the 1970s; now those patents are considered the foundations of the biotechnology industry.

Acknowledging that no one at this point seems to know what to make of the NIH application, NIH director Bernadine Healy announced that the next move would be made by her administrative superiors; Louis Sullivan, secretary of the department of Health and Human Services, which runs NIH, will make a policy decision on filing future cDNA patents. But before he does, the White House will be asked for an opinion. Healy said that the White House Office of Management and Budget, the Office of Science and Technology Policy and possibly the Council on Competitiveness will all participate in making gene-patenting policy.

Continuing controversy on this issue may not bode well for future harmony in the genome project, but it does appear to take the immediate pressure off NIH. When the news of NIH's application first broke last month, many researchers blamed both the agency and Craig Venter, the NIH scientist who discovered the genes, for filing the application in the first place. Now that industry and legal experts have turned out to be divided on the issue themselves, most of the participants agreed that

NIH had been right all along in taking the safe route: given the uncertainty in the viability and implications of gene patents, letting the sequences appear in the public domain without patent protection would have been tantamount to legal "malpractice", said patent expert Harold Wegner of George Washington University.

Given the lack of consensus on the merits of gene patents — or even agreement on the application of standard patent law to this case — some lawyers at the conference argued for special patent-of-office treatment for gene patent applications until the issue is settled. S. Leslie Misrock, a partner in the patent law firm Pennie and Edmonds, proposed that the patent office accept cDNA applications on a special reduced-fee basis until it had decided whether they could be granted. If cDNA patents are approved in principle, Misrock suggested, the patent office could allow each applicant the opportunity to pay the full fee and have his application examined for approval as usual.

Lawyers at the conference noted that the US Congress has recently raised another issue that appears to bear on the controversy: that of the extent to which scientists are entitled to use patented materials in the course of their research, without paying royalties. Long-standing patent law allows for free use if the researcher has only "philosophical interest" in the patented material. But in a bill introduced last year, the House judiciary committee questioned the meaning of that wording in modern laboratories, where researchers are encouraged to turn every invention into a marketable product. The committee plans to hold a hearing on the issue before the end of the year.

Christopher Anderson

A battle as old as the nation

Washington

BEHIND the current gene patent controversy lies a debate that has raged since Thomas Jefferson introduced the first US intellectual property laws nearly two hundred years ago: When are patents good for industry and investment, and when do they just get in the way?

That question is especially contentious in research, where patent procedures often clash with ideals of free exchange of scientific data. US patent applications are by law secret and can take years to be approved; during that period applicants may refuse to publish or share the data for fear of losing all rights if the patent is not granted.

Until a 1980 Supreme Court decision allowing the patenting of oil-eating bacteria, the United States took the moral high ground: it put all government research (such as that done by the National Institutes of Health (NIH) and the

laboratories of the Department of Energy) in the public domain. This was despite some very good evidence that the policy was doing the US taxpayer few favours — a study conducted in the 1960s found that no government invention to date had ever been commercialized. During World War II, the government itself had to develop penicillin to combat infections in troops because no company would invest the money developing an antibiotic it could not own.

Universities also wrestle with balancing science and commerce. At a technology transfer meeting at NIH last week, Hoffman-LaRoche patent attorney George Gould noted that until the 1970s Harvard University had a policy of never seeking patents for the inventions of its scientists. "They changed their mind when they realized that 100 years of science from the best minds in medicine had never turned out a product," Gould recalled.

Although most US research universities now have technology transfer offices to encourage patents, a new spectre — product liability — has arrived to cloud the issue. Reid Adler, NIH's technology transfer director, reported that increasingly aggressive litigation to recover damages in medical-product liability cases is starting to raise fears throughout the patent chain. If a university licenses its technology to a small company to manufacture a product that eventually becomes the subject of a lawsuit, a plaintiff may choose to sue the university — "going for the endowment", as Adler puts it — as well as the company. In the face of such prospects, some universities are starting to reconsider the benefits of technology transfer, especially to start-up companies. "They're saying, 'We'd like to license the technology, but we're afraid of losing the football team,'" Adler said. **C.A.**

African rift in Kyoto

London

NEXT year's meeting of the Convention on International Trade in Endangered Species (CITES) promises to be a stormy affair, now that a CITES-appointed expert panel has endorsed a South African plan to begin a controlled trade in ivory from that country. If the 112 nations that have signed the convention agree with the panel's assessment when they meet next March in Kyoto, Japan, the international trade in ivory will be reopened, less than four years after the CITES signatories decided that a complete ban was needed to preserve Africa's rapidly dwindling elephant populations (see *Nature* 341, 555; 1989).

Many observers expect South Africa's request to be rejected, but the debate surrounding the plan — and, more significantly, similar requests from five other southern African countries now being examined by CITES experts (see *Nature* 351, 7; 2 May 1991) — will bring to the surface deep divisions among the African nations over how best to conserve their elephant populations.

In one camp stand the southern African nations requesting that their elephant populations be removed from Appendix 1 of CITES, which lists species (and populations within species) in which no trade is permitted. This group, comprising Botswana, Zimbabwe, Malawi, Namibia and Zambia, points out that its elephant populations are not endangered, and argues that the revenue gained from sales of ivory can help conservation by financing anti-poaching patrols and providing economic rewards for local people (who otherwise have no good reason to support elephant conservation, as the animals often compete for valuable agricultural land).

Opposing them is another group, including Kenya, Tanzania, Ghana and Zaire, which admit that they do not have the resources to control poaching, and so want the ivory trade to stay banned until Western nations provide the necessary financial aid for them to manage their elephant populations effectively. The expected clash between the two groups in Kyoto is acutely embarrassing to black African governments, which like to show a united front in international negotiations involving the developed world (although politically isolated South Africa is not bothered by it at all).

In 1989, when all African elephant populations were moved onto Appendix 1, most CITES signatories decided that the trade had to be banned entirely, despite the reasonable claim from several southern African countries that

their wildlife management practices were sufficiently sound to allow a controlled trade without endangering their elephant populations. Given that many other African countries were unable to stamp out serious problems with poaching, the 1989 CITES meeting in Lausanne, Switzerland, decided that allowing any legal trade would provide a cover through which illegal traders could 'launder' their poached ivory. The decision came after a decade in which the total population of African elephants dropped from 1.3 million to a mere 650,000.

This approach has been successful, now that the two main importers of ivory, Japan and Hong Kong (represented within CITES by Britain), have accepted the CITES decision, slashing the demand for ivory. The trade has withered, and poaching has been reduced — by a factor of almost 100 in Kenya, for example.

Many conservationists and ecologists believe the same objections to a legal ivory trade that convinced the CITES signatories in 1989 also hold today, and say that the ban's success should not be jeopardized by its relaxation (see *Nature* 351, 265; 23 May 1991). So why did the five-member CITES panel disagree, when confronted with the South African proposal?

One important factor is the advent of two new techniques to identify the origin of particular tusks. Before the trade was banned, tusks were often identified only by serial numbers written in supposedly permanent felt-tip pen. Not surprisingly, poachers found it relatively easy to disguise the source of their tusks. But the new South African plan includes the marking of tusks with tamper-proof adhesive holograms, developed by the South African Council for Scientific and Industrial Research; the holograms disintegrate when removed and will be extremely difficult to forge. The plan also relies on a new technique, first described in *Nature* (346, 744; 1990), which measures the isotope ratios of carbon, nitrogen and strontium in ivory

samples, and which can be used to determine the particular region from which an individual tusk originated.

In its report, released earlier this month, the CITES panel says that these new developments will reduce the chances of laundering illegal ivory through a CITES-sanctioned South African trade and concludes that South Africa's elephants could be moved onto Appendix 2 of CITES, allowing a carefully controlled trade in South African ivory, without increasing the threat of poaching.

But despite the CITES panel's verdict, the plans to resume a limited trade can expect a rough ride in Kyoto, both from the African states that oppose their neighbours' approach and from many developed nations. Steve Cobb, from the Environment and Development Group in Oxford, which has helped many of the African nations to produce elephant conservation plans, points out that the stable wildlife management policies in those countries now asking for a resumption of the ivory trade may not last. In Zimbabwe, for example, he expects the importance of conservation in land-use policy to "drop below the horizon" in the coming years as a brewing row over land ownership comes to a head: more than a decade after Robert Mugabe came to power, little of Zimbabwe's land has come into the hands of the black majority. Also, many of the African states with which Cobb has been collaborating want to use his conservation plans to attract aid from agencies such as the World Bank. Until these deals have been negotiated, most of these countries want the ivory trade to stay banned.

Most conservation experts are unwilling to forecast the outcome of Kyoto meeting. "Who can predict what will happen when you put 112 countries in one room?" asks Jonathan Barzdo, a member of the panel that examined the South African plan. But if the southern African states fail in their attempt to obtain CITES approval to trade in ivory, they still have the option to try to sell their ivory anyway. All six nations entered a reservation to the Appendix 1 listing after the 1989 meeting.

Although none of the southern African states are trading in ivory at the moment, the reservations enable them to ignore the CITES ban. The only problem will be finding importers for their tusks. "Nothing has been decided," says Herman Grove, from the South African Department of Environment Affairs, "but the possibility for trade is still fairly limited." Japan and Hong Kong have no reservation against the ivory trade ban, and it would take several years for other countries that have not signed CITES to expand their ivory working industries and create a significant demand.

Peter Aldhous



Will removing the ivory ban lead to more of this?

South Korea sets sights on G-7 status

Seoul

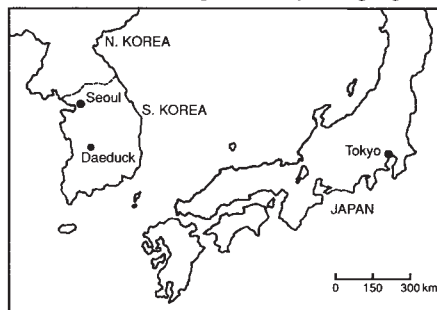
WITH commercial achievements that parallel Japan's success in the 1970s, South Korea has already achieved a reputation as one of Southeast Asia's economic "little dragons", along with Taiwan, Hong Kong and Singapore. Now the country hopes to follow in Japan's footsteps by establishing itself as a scientific and technological power.

To do that, the South Korean government is planning to launch a \$7,000-million research-and-development project next year with the goal of bringing the country's technological capabilities up to the level of the world's leading industrial nations by 2000. Named the 'G-7 project' after the G-7 group of Western summit nations (the United States, Japan, Germany, France, Italy, Canada and the United Kingdom), the project would be by far the largest single investment ever made in science and technology by the Korean government.

Before the project can get under way officially, the Korean parliament will have to pass the G-7 budget request for fiscal year 1992 (which begins on 1 January), a move that should come in the next few

weeks. Government researchers say that science-related budgets undergo only minor modification during the parliamentary process, so the project is certain to go ahead. The biggest hurdle was overcome earlier this year when the Ministry of Science and Technology and other science-related ministries, such as the Ministry of Trade and Industry, decided to launch the project in 1992.

South Korea hopes that by bringing the



level of its science and technology up, it will be able to join the powerful G-7 group early next century. The political process of democratization in South Korea from 1987 has dramatically increased wage levels and the demand for welfare, pushing

up the cost of production and reducing the international competitiveness of Korean export products. Furthermore, heightened protectionism of technology by advanced nations is restricting the advance of Korean industry, and the only way forward is for the country to catch up with the advanced nations by developing its own indigenous high technology to make value-added products for export. The target of catching up by 2000 may seem ambitious, but over the past three decades South Korea has shown a remarkable ability to develop its science and technology infrastructure.

With the partitioning of Korea at the 38th parallel at the end of the Korean war in the early 1950s, the southern half of the country was left with a largely agricultural economy, few natural resources, a meagre industrial base and very few facilities for research and development. In 1965, government and industry spending on research and development amounted to less than 0.3 per cent of the nation's then-tiny gross national product (GNP).

But as a result of a concerted effort by government and industry, annual investment in science and technology now stands

White lab coats and greasy hands

Seoul & Daeduck

THE quest to secure economic competitiveness has pushed South Korea to establish a number of scientific and technical institutes. Most recently, it has decided that more of its researchers need to get their "hands greasy" if the country is to keep its edge.

The Korea Institute of Science and Technology (KIST) opened in Seoul in 1996 to carry out contract research for industry. By the 1980s its industrial contracts had diminished considerably, as large-scale industries in Korea began to do their own research, so KIST concentrated on longer-term research of national interest. In 1981, it merged with the Korea Advanced Institute of Science (KAIS), a graduate school for scientists and engineers established in 1971 to attract Korean scientists back from the United States and Europe, and to keep the most promising Korean students in the country.

The merged institution, the Korea Advanced Institute of Science and Technology (KAIST), has become a successful university based in the new science town of Daeduck, south of the capital. Its faculty has grown from 30 members in the 1970s to more than 300 now, and it currently turns out each year about 700 doctoral and masters scientists and engineers — 20 to 30 per cent

of South Korea's total output.

KIST — which in 1989 once again became independent from KAIST — has been equally successful, but in a different way. Several new institutes have been created from it, including the Korean Ocean Research and Development Institute, the Centre for Science and Technology Policy, the Systems Engineering Research Institute and the Genetic Engineering Research Institute in Daeduck.

There have been, however, a number of criticisms that KIST and KAIST are not serving the immediate interests of industry, and, partly in response to these complaints, the Korean government established the Korea Academy of Industrial Technology (KAITECH) in 1989. The academy, which currently receives government funds of about \$200 million a year, exemplifies an ongoing change in government science policy: ministries other than the Ministry of Science and Technology — which until now has had the role of being the lead science and technology ministry — are getting deeply involved in research and development policy.

"For twenty years we have tried to adopt the American way of doing research by setting up high-class academic research organizations like the Korea Institute of Science and Technology," says KAITECH president Young Wook Kim, but what is needed, as Japan has shown, is "produc-

tion engineering".

Kim complains that "80 per cent of PhDs in Korea don't contribute to industry. They just play with chalk on blackboards." Instead, he would like scientists to "take off their white coats, put on some blue overalls, and get down among the grease".

KAITECH is offering powerful financial incentives to entice Korea's PhDs to get their hands greasy. The academy will provide 70 per cent of funds, in the form of zero-interest loans, for near-market joint research by academics and industry on such projects as high-definition television. The remaining 30 per cent will come from the companies involved in the research.

If the research is successful in producing a commercial product, the company (or companies) has to pay back 70 per cent of the research funds received from the government. An unusual characteristic of KAITECH support is that forty per cent of these returned funds will go to the individual researchers involved in the successful research project in the form of grants to support the research of their choice.

The idea is to get academics involved in near-market research by offering them the incentive of obtaining research funds from industry for their own academic pursuits.

David Swinbanks

A science tax?

Seoul

THE world's scientists, who often feel they do not have enough money to support their research, will certainly appreciate an idea making the rounds in the South Korean government — a tax levied specifically to raise research funds for science and technology.

As Korea tries to catch up with the technology of the advanced nations of the world, government bureaucrats are giving serious consideration to the introduction of a 'science tax'. Although such a tax is unprecedented, the general scheme of assessing taxes for specific purposes is well established — in the past, the South Korean govern-

ment has put extra taxes on corporations and its citizens to raise funds for defence. So why not, the bureaucrats ask, have a tax to aid Korea's bid to catch up with the technology of Japan and the West?

Among those monitoring science policy in Seoul, the science tax is the talk of the town, but no one can say whether it will ever be levied. It is bound to be unpopular with the general public, and government researchers say that one thing is certain: no attempt to introduce the new tax will be made until, at the earliest, after the general election for the national assembly early next year.

David Swinbanks

at about 2 per cent of a healthy GNP, which the government officially placed at about \$273,000 million for 1991. And last week, in mapping out Korea's seventh five-year economic plan, for 1992-1996, a committee chaired by Prime Minister Chung Won-shik estimated that science and technology investment will rise to 3.2 per cent of GNP, or 11.5 million million won (\$15,300 million) by 1996.

These figures should be treated with some caution because much of the increase is due to a surge in capital spending rather than to increased operating expenses for research. Private conglomerates such as Hyundai and Samsung have been funding the construction of research institutes, and this construction boom is expected to continue into the 1990s. Also, the government consistently tends to underestimate GNP in order to maintain favourable lending rates with the World Bank.

Nevertheless, South Korea does now have a substantial infrastructure of new and well-equipped research institutes in both the government and private sector, and the population of doctoral- and masters-level engineers and scientists is growing rapidly. At present, the country has about 50,000 researchers (one-tenth that of Japan), and the government's plans call for this number to be tripled by early next century. This would bring the percentage of researchers in the total population (currently 42 million) up to a level comparable with that of advanced nations, such as the United States and Japan.

The G-7 project forms a key part of the government plans for expansion. With a proposed budget of about \$7,000 million over nine years (1992-2000), the project will support the development of seven advanced technologies — semiconductor memory chips, Integrated Services and Data Networks (ISDN) for telecommunications, high-definition television (HDTV), electrical vehicles, intelligent computers, new antibiotics and chemicals for agriculture, and advanced, fully auto-

mated manufacturing systems.

A second part of the project will concentrate on the development of basic technology that is further from the market. Again seven areas are targeted: new materials, next-generation transportation systems, biotechnology, 'environment friendly' technology, new clean energy sources, advanced atomic reactors and human interface technology based on 'user friendly' electronics and robotics.

If all goes according to plan, Korea will have developed 256-megabit dynamic random access memory (DRAM) chips by 1996 and 1-gigabit chips by 2000. (Korea already has its own domestically developed 16-megabit chips.) HDTV monitors compatible with Japanese and European systems are scheduled to become available in 1993, and HDTV flatscreen displays by 1997. Electrical vehicles would be commercialized by 1996, a neuro-computer by 1997, and ISDN by 2000.

Much (50 to 70 per cent) of the research will be carried out by institutes affiliated with the Ministry of Science and Technology, particularly the Korean Institute of Science and Technology (KIST) in Seoul and affiliated KIST institutes and the Korean Advanced Institute of Science and Technology in Daeduck science town. But private industry will also make a substantial contribution, and so will the new Korean Academy of Industrial Technology, an organization that coordinates and funds near-market joint research between government, industry and academia.

Whether the project will achieve the full level of funding proposed remains open to question. One Western science officer in Seoul says he always divides figures quoted by Korean government officials by three to arrive at a more realistic estimate of what is actually going on. It is clear, however, that the G-7 project will go ahead and Korea is determined to go on the fast track to try to catch up with the most advanced nations of the world.

David Swinbanks

UK primate research under scrutiny

London

A NUMBER of published British research studies involving primates should not have been licensed under existing animal welfare legislation, according to the Scottish pressure group Advocates for Animals. The group, which last year exposed serious breaches of animal welfare legislation at the National Institute for Medical Research of the Medical Research Council (MRC) (see *Nature* 345; 190; 1990), this week released a report detailing 13 research papers published since 1987 derived from work that Advocates for Animals believe should not have received Home Office approval.

Les Ward, director of Advocates for Animals, said the aim of the report is to ensure that British primate researchers adhere to UK animal experimentation legislation, as overhauled in 1986. The 1986 Act requires that the Home Secretary not grant a project licence for work involving primates unless no other species are suitable for the research, or if it is impractical to obtain animals from other species.

Ward questions whether this requirement was met for several of the research projects examined in the report. Advocates for Animals also charge that some of the research was repetitive of previous experiments, and that in some cases the Home Office failed adequately to balance the likely suffering of the animals concerned against the expected benefits of the research. The group wants the Home Secretary to investigate why the named research projects were approved, and for the Animal Procedures Committee (a panel of independent experts which advises the Home Office) to review all current British primate research. On Monday, the Home Office said it would look into Advocates for Animals' specific allegations before deciding whether to examine British primate research in general.

Unlike its previous report on work at the National Institute for Medical Research, which involved clear breaches of the 1986 Act, Advocates for Animals' latest charges will be difficult to address. The issues in question — whether the use of primates was essential, whether the suffering outweighed the gain in scientific knowledge, and the degree of overlap with previous work — depend on subjective judgements, where the opinions of researchers and of animal welfare groups will diverge. "You can't give somebody a cost-benefit equation in which to plug in the parameters," says MRC secretary Dai Rees, whose organization provided funding for more than half of the projects named by Advocates for Animals. Nevertheless, he says the MRC intends to investigate the charges fully. Peter Aldhous

Caught in a propaganda war

London

CROATIA'S leading research centre has been dragged into the bitter conflict between the breakaway republic and the Serbian-dominated Yugoslav government. The federal government's Tanjug press agency and Belgrade's television station and leading newspapers have accused the Ruder Bošković Institute in Zagreb of working secretly to produce an atomic weapon for use against Serbia. Western physicists familiar with the Ruder Bošković Institute dismiss the claim as impossible, but its director-general, Krunoslav Pisk, fears that the propaganda offensive is a prelude to an aerial bombardment of his institute.

The accusations, which surfaced in the Belgrade media last month, implicate the Ruder Bošković Institute in an alleged Croatian plot to produce a nuclear weapon with help from South Africa. But Pisk protests that his institute is devoted entirely to peaceful activities, including theoretical physics, radioactive waste disposal and food irradiation. In any case, he says, the Bošković institute lacks the high-technology hardware needed for weapons research and production. The institute's nuclear equipment is limited to a cobalt source, a Van de Graaff generator and an ageing accelerator, Pisk says. A cyclotron and a neutron source are now defunct, he

adds.

Researchers in the West back Pisk's description of the institute. "It's absolutely ridiculous" to suggest that the Bošković institute could be involved in a weapons project, says Edward Bilpush, a nuclear physicist from Duke University in North Carolina who has visited Zagreb on many occasions. "I have complete access whenever I go there."

Walter Greiner, a theoretical physicist at the University of Frankfurt who last visited the institute in May, shortly before Croatia's secession from Yugoslavia, says that the Bošković institute's experimental physics equipment is too primitive to support an atomic weapons programme (the institute's main strength is in theoretical physics).

Pisk's fear is that the accusations made in the Belgrade media are part of a campaign to justify a planned military attack against his institute — which has so far escaped unscathed from the limited air strikes against Zagreb. The Bošković institute is Croatia's only research centre of international standing, and Pisk estimates that it represents 60 to 70 per cent of the entire scientific infrastructure of the republic. Pisk believes that the federal government wishes to destroy Croatia's infrastructure so that the republic can no longer

function independently of the federal government.

In an attempt to counter the claims emanating from Belgrade, Pisk has written to Hans Blix, director-general of the International Atomic Energy Agency (IAEA) in Vienna, saying that the institute would welcome a team of IAEA inspectors. But Blix's hands are tied by IAEA's constitution, which says that he can respond only to complaints made by IAEA members. As the breakaway former Yugoslav republics are not represented within IAEA, Blix has been able only to contact the Yugoslavian foreign ministry saying that IAEA is willing to inspect the Bošković institute, should the Belgrade government request it to do so. Not surprisingly, the federal government has made no such request.

That the Bošković institute should become embroiled in a conflict hinging on opposing nationalist ambitions is sadly ironic. Bilpush says that for many years the institute "was a gateway to the West for the East", bringing together researchers who were otherwise separated by the political division of Europe. Both Western and Eastern bloc scientists could travel freely to communist, but non-aligned, Yugoslavia, so the institutes' researchers were able to organize regular international meetings in Zagreb and in other cities throughout Yugoslavia.

Peter Aldhous

A chip off some old rocks

Washington

THROUGH a combination of ingenuity and a little luck, mission scientists at the Jet Propulsion Laboratory in Pasadena, California, were able last week to bring down to Earth a few frames of the many pictures taken by the spacecraft Galileo as it passed within 10,000 miles of the asteroid Gaspra. Galileo, due to arrive at Jupiter in 1995, is in working order except that its high-gain antenna failed to open, limiting its data transmission to a less powerful telemetry link that transmits at the meagre rate of 40 bits per second.

It was originally intended that the pictures of Gaspra, stored on tape aboard the spacecraft, would not be seen until late next year when Galileo is due to pass close by the Earth to gain some final momentum before swinging off into the outer solar system. But project scientists realized soon after the 20 October encounter with Gaspra that their estimate of its position relative to the spacecraft was more accurate than they had expected, so that the asteroid was fully imaged in the central frame of a three-by-three mosaic taken to

ensure adequate sky coverage. After a test transmission, they were able to determine which section of the taped data contained the complete image. Even so, it took about 20 hours to receive this one image.

Gaspra is a chunk of rock about 15 kilometres across at the broadest part. It resides in a section of the asteroid belt, beyond Mars, where the expected lifetime between collisions is some hundreds of millions of years. The amount of cratering visible is consistent with such an age, and the irregular shape lends weight to the idea that asteroids are fragments created by the collision and disruption of larger

parent bodies that formed at the same time as the planets.

The complete set of pictures taken by Galileo includes images at different wavelengths and at higher resolution. When all of the data have been transmitted to Earth, it will be possible to create a true colour picture of the asteroid, which may help in determining its surface structure and composition. But that will have to wait. As Galileo recedes from the Earth, the transmission rate will fall to a paltry 10 bits per second, and will not come back up to 40 bits per second until next April, as it nears the Earth *en route* to its final swing towards Jupiter. In the meantime, mission scientists will try once again to release the jammed antenna by switching off some of the spacecraft's instrumentation and rotating the antenna side away from the Sun for two days. The drop in temperature, it is hoped, will dislodge alignment pins that are keeping the ribs of the antenna from opening out. Previous cooling manoeuvres in July and August failed to free the antenna, but mission managers believe that they did cause the pins to "walk towards freedom".

David Lindley



PCR DNA typing for forensics

SIR — The first admission by a British court of forensic evidence using the polymerase chain reaction (PCR)^{1,2} suggests that acceptance of this test could quickly become general outside the United States³. In actual casework, great care must be exercised in the areas of chain integrity (continuity), quality control and the choice of the population database used in the probability calculations. Otherwise, unreliable conclusions may be drawn and the expert witnesses accused of presenting worthless evidence. The PCR procedure has had only limited forensic application, but in individual cases it has allowed DNA analysis otherwise inaccessible to standard procedures. This is so when the amount of DNA is limited, or when the DNA is degraded because of environmental factors, contains a mixture of DNA coming from various subjects or is contaminated by handling before reaching the laboratory.

Yet PCR analysis was admitted by the Italian courts in April 1991 (Corte d'Assise, L'Aquila) as part of the trial evidence in the murder of a seven-year-old girl (the Balsorano-Avezzano murder, 1990). On 23 August 1990, C. C. was found killed in a wooded area near the family house. The police investigation pointed to the victim's maternal uncle as the suspect. The major evidence consisted of hair roots and blood spotting found on the accused's vest and underpants.

The investigators had DNA prints prepared from the victim's and the accused's blood. The victim's profile matched those from the hair-root cells and bloodstains. The DNA was examined by *in vitro* amplification using the PCR of different polymorphic sites, including VNTRs loci (APOB, D1S80, HLADQalpha, D2S44), microsatellite markers (APOCII, 3'CACA distrophin, CA/GTVNDRI8) and two allele site specific RFLPs (XV2C, KM19, CKMM, APOCII/BanII/AvaII). In addition, a 49 bp portion from a 3.4-kb Y chromosome-specific tandem repeat was amplified for sex determination of the biological traces, and provided no evidence of specific Y-material.

The PCR results determined that DNA profiles from the victim's blood specimen and traces matched with 1.7×10^{-11} :1 odds against the match arising by chance in the population. This forensic case was also investigated by direct sequencing of the CA/GT strand of the amplified microsatellite VNDRI8. The victim's blood pattern matched that of hair-root cells found on the uncle's vest. The suspect received a life sentence.

It has been argued that the DNA

typing of small-sized forensic samples should be based on different means, including the determination of amplified DNA fragment length differences, hybridization with allele-specific oligonucleotide probes, and direct DNA sequencing⁴. The 1990 Balsorano-Avezzano murder proves that this complex protocol provides reliable, probative, not unduly prejudicial, and helpful biological evidence to the jury.

BRUNO DALLAPICCOLA
GIUSEPPE NOVELLI

Department of Human Genetics,
IInd University of Rome,
Via Ramazzini 15, 00151 Rome, Italy

ALDO SPINELLA

Central Criminal Police,
III Scientific Division,
Viale Aeronautica 7, 00144 Rome, Italy

1. Davies, K. *Nature* **352**, 381 (1991).
2. Hagelberg, E. et al. *Nature* **352**, 427–429 (1991).
3. Kirby, L. T. *DNA Fingerprinting* (Stockton Press, New York, 1990).
4. Higuchi, R. et al. *Nature* **332**, 543–546 (1988).

Chauvinism

SIR — A recent study tends to support the hypothesis of 'national publication bias', according to which journals prefer to publish papers from their own country¹. This may be true for the journals published in the West but is certainly not true for those published elsewhere. There is also the problem of the 'impact factor'².

Most of the scientific journals published in India or other developing countries do suffer simply because of a pre-conceived notion that the large number of papers published in them have a low impact factor rating. This may or may not be true. Fortunately, these journals are striving hard to improve their image by internationalizing their editorial boards to increase what is called their 'internationality factor'³. As national publication bias and impact factor are to some extent interrelated, this could very adversely affect scientists, especially in developing countries. This fact should be of some concern to the whole scientific community.

As professional career decisions and even election to fellowships of the national academies in developing countries also depend on impact factor rating, as in the West, scientists from these countries must publish in journals with a high impact factor rating, which may or may not be a true reflection of the capability of a scientist.

Nearly all the journals with high impact factor ratings are published in the West and, because they suffer from national publication bias, authors from

developing countries are at a disadvantage; a very vicious circle indeed. In my view, university authorities and scientific academies should think of ways to minimize the harmful effects of national publication bias and low impact factor. Editors of journals with a high impact factor rating may also have to change some of their policies to ensure that science remains free of national bias and is truly free and international in outlook.

K. N. MEHROTRA

Division of Entomology,
Indian Agricultural Research Institute,
New Delhi 110 012, India

1. Ernst, E. & Kienbacher, T. *Nature* **352**, 560 (1991).
2. Garfield, E. *Science* **178**, 471–479 (1972).
3. Sen, B. K., Karanjai, A. & Munshi, U. M. *Indian J. Malariol.* **27**, 197–204 (1990).

Modern myth

SIR — You remark: "Perhaps biologists... look forward to the time when interesting questions can be answered" (*Nature* **352**, 11; 1991), in relation to Walter Gilbert's article (*Nature* **349**, 99; 1991)², which includes the astonishing statement: "Thus I expect that... half of the total knowledge of the human organism will be available in five to seven years, and all of it by the end of the decade" (*Nature* **349**, 99; 1991). But it is not information we should seek, but understanding.

It is a modern myth that the DNA of an organism contains all the information necessary to produce that organism. This is not so. Some components of the body, such as the phospholipid bilayer and the structure of the mitochondrion, do not seem to be coded by genes, but are built onto structures transmitted in the cytoplasm of the ovum. Widely different morphologies can be produced from similar genomes, such as some castes of social insects, which depend on the physiological regime within which the genome is expressed. Similarly, correct expression of the information in the human genome depends on many aspects of the external environment and on all the myriad micro-environments created by the contortions of morphogenesis, to which multiple copies of the same genome become exposed in the cells of the embryo. The number of different expressions of one genome is potentially possibly infinite.

Although the complete sequencing of the human genome will be a very major achievement, with profound implications for the future health of mankind, it will fall very far short of making available all there is to be known about the human organism.

DORIAN PRITCHARD

Division of Human Genetics,
University of Newcastle upon Tyne,
Newcastle upon Tyne NE2 4AA, UK

Czechoslovak environment

SIR — Western observers shocked by the extent of environmental damage in Czechoslovakia^{1,2} may not be aware of the close relationship between democracy and environmental protection. In Western democracies, the environment is protected in the interplay of government, nongovernmental organizations and unbiased science. In the former Soviet Union and Eastern Europe, science has been deformed as a result of previous government control and marxist ideology. Environmental science was almost neglected in Czechoslovakia; the Academy of Sciences, for example, had no environmental programme.

Even so, several institutes and universities studied the environmental degradation on a limited scale. Their results were usually not made publicly available; their dissemination was partly prohibited by law⁴. These studies are summarized in two reviews^{3,4} issued after the political change in 1989: the 'federal' report³ written in 1989 by a team of scientists and environmentalists, and the 'Czech' report⁴ written in 1990 by a team of the Czech Ministry of Environment.

The number of references in corresponding chapters of both texts is limited as well as their availability (see table). The references are not properly marked in the text, which indicates poor practice in using scientific information. The small number of primary sources cited in the federal report is obviously because of their limited availability in 1989. We have experienced this problem in reviewing the use and fate of polychlorinated biphenyls in Czechoslovakia⁵. The number of references given in the Czech report, which was completed after the political change, is, however, even lower (see table), and parts of the text are identical with the federal report, for example, 80 per cent of the text of the chapter "Living Nature".

By using the *Science Citation Index* (SCI), one can guess the publication activity of the authors of the federal report and the authors of citations used in both reports (see table). A group of 50 environmental scientists was created randomly for a comparison using the contents of recent issues of seven international journals (*Chemosphere*, *Environ. Intern.*, *Environ. Pollut.*, *Intern. Environ. Study*, *J. appl. Ecol.*, *J. exp. Marine Biol.*, *Ecol.* and *J. Freshwater Ecol.*). According to the SCI, 1990, 67 per cent of the international group is mentioned in papers of other scientists compared with 25 per cent of the authors of the federal report.

Both reports contain primary data on human health, air, water and soil pollu-

tion that are unavailable elsewhere. As the measuring techniques are not mentioned, it is difficult to compare these data with results of others. A better situation seems to obtain in the biological sciences ("Living Nature") where authors could publish their studies. It can be concluded that normal scientific communication and processing of information has not been developed in Czechoslovak environmental science. This may

inhibit the public understanding of the environmental situation in the country.

LUBOS NONDEK

*Water Research Institute,
Podbabska 30,
16000 Prague 6,
Czechoslovakia*

1. French, H. F. *State of the World 1991* (ed. Brown, L. R.) Ch. 6 (W. W. Norton, New York, 1991).
2. Hamilton, J. M. *Bull. Atom. Scientists* **47**, 12 (1991).
3. Vavrousek, J. et al. *State and Trends of the Environment in Czechoslovakia*, 3rd Edition (Cesky svaz ochrancu prirody, Praha, 1990, in Czech).
4. Moldan, B. et al. *Environment of the Czech Republic* Academia Praha 1990, (in Czech).
5. Nondek, L. & Frolikova, N. *Chemosphere* (in the press).

NUMBERS OF RELEVANT AND AVAILABLE REFERENCES AND THE CITATION OF THEIR AUTHORS IN SCIENTIFIC LITERATURE

Chapter*	Czech report (1990)			Federal report (1989)		
	Ref.†	Avail.§	Cit.(%)	Ref.†	Avail.§	Cit.(%)
Air	6	1	0	7	1	0
Water	5	0	0	9	3	0
Soil	9	2	0	12	2	0
Living nature	15	9	70	15	9	70
Xenobiotics†	5	2	60	5	2	60
Forestry	11	4	15	33	9	30

* See the titles in English edition of the Czech report issued by Czech Ministry of environmental protection without (sic!) Tables and Figures; the compared parts represent 26 per cent (the Czech report) and 77 per cent (Federal report) of the text. † 'Autochthonous Substances' in English edition. ‡ Only sources relevant to Czechoslovak environment accounted. § All books and periodic journals are supposed to be publicly available.

Brownian motion

SIR — Recently, in an abstract¹, Daniel H. Deutsch has argued that Robert Brown^{2,3} could not have seen the random motion of small particles, later called brownian motion. Deutsch's main arguments are that Brown's system was too noisy, that he did not use coverslips (not yet invented), that his particles were too large (for example pollen grains), too light or too heavy, and that a proper achromatic microscope had not yet been invented. Deutsch says that to see real brownian motion in water at magnifications of 350× requires particles of approximately 1 micrometre and a rigid system free from vibration and evaporation.

To learn what Brown really saw in 1827, there are two requirements; a careful reading of Brown's publications and use of his (mainly single lens) microscopes. Both courses have been followed by Brian Ford⁴, who gives a vivid description of Brown's findings as well as the discovery and restoration of one of Brown's microscopes, at present at the Linnean Society in London. It is not clear that Deutsch used these methods. If he had read Brown's papers more thoroughly, he would have noted that Brown did not describe the movements of pollen grains, which indeed are too large for Brownian motion, but described movements of particles *inside* pollen grains (see also the title of his paper). Brown estimate the size of these particles as 1/15,000 to 1/20,000 of an inch (1.7–1.3 µm). Later, he added that, using other single lens microscopes as

well as the best achromatic compound microscopes available, he could see movements of even smaller particles (down to 1/30,000 of an inch = 0.85 µm).

A careful reading of Brown's observations refutes most of Deutsch's arguments. The particles were not too large, but of the required size of around 1 µm. Brown makes sure that the motions were not due to currents in the fluid nor to its gradual evaporation, a problem he was well aware of.

To diminish currents due to evaporation, Brown even immersed small droplets of water, containing his microscopic particles, in almond oil. The 'brownian' motion remained visible. He also mentions that, in some grasses, the membrane of the pollen was so transparent that motion of the particles could be seen inside the intact pollen grains, where currents or evaporation can be excluded as a cause of the movements observed. Our conclusion, therefore, must be that Brown did see real Brownian motion.

GERHARD C. CADÉE

*Netherlands Institute for Sea Research,
PO Box 59, 1790 AB Texel,
The Netherlands*

1. Deutsch, D.H. *Bull. Amer. Phys. Soc.* **36**, 1374 (1991).
2. Brown, R. *A brief account of microscopical observations made in the months of June, July and August, 1827, on the particles contained in the pollen of plants and the general existence of active molecules* (printed privately, London, 1828).
3. Brown, R. *Additional remarks on active molecules* (printed privately, London, 1829). (Both papers reprinted in *The Miscellaneous Botanical Works of Robert Brown* (ed. Bennett, J.J.) **1**, 463–486 (Roy Soc. London, 1866).
4. Ford, B.J. *Single lens, the story of the simple microscope* (Harper & Row, New York, 1985).

Methane on the greenhouse agenda

Kathleen B. Hogan, John S. Hoffman and Anne M. Thompson

Methane is often ignored in the greenhouse debate. Reductions of emissions of this gas would be relatively easy to achieve and could have a significant effect on global warming.

IN February 1991, the United States government proposed that a comprehensive framework for greenhouse gases would be preferable to focusing on a single gas as a strategy to limit greenhouse warming. We report on recent results on opportunities for reducing emissions of methane, a gas frequently overlooked in the greenhouse debate.

Methane concentrations are rising at about 1 per cent per year (10–16 parts per 10^9 by volume) and similar increases are predicted^{1–3}. Increases in methane are largely correlated with increasing populations, about 60 per cent of global methane emissions being associated with human activities (Fig. 1). Relatively small reductions in methane emissions of 40–60 Tg per year (about 10 per cent of annual emissions) would halt the rise, assuming that the rate of methane destruction by OH remains the same^{2–4}.

Approaches to reduce overall emissions of greenhouse gases will be enhanced through the design of effective programmes to reduce methane emissions for several reasons. First, methane is a potent greenhouse gas and reductions in emissions would be 20–60 times more effective in reducing the potential warming of the Earth's atmosphere over the next century than reductions in CO_2 emissions². Second, methane released by human activities can generally be viewed as a wasted resource, opening the possibility for reductions to be low cost, if not profitable. Third, reductions in methane emissions will provide additional benefits of reducing the risks of increasing tropospheric ozone and reducing the

Earth's oxidizing potential^{3,5–8}. Model calculations show that stabilizing methane concentrations could substantially reduce the expected global tropospheric ozone increase between 1985 and 2100 and, although more uncertain, that OH would return to present-day levels by 2100 after a significant suppression⁹.

Because of the potency of methane in the atmosphere and its relatively short lifetime, stabilizing methane concentrations will have a substantial impact on reducing potential warming. The results of holding methane emissions at about constant levels (540 Tg) throughout the next century are shown in Fig. 2 (using projections from the Intergovernmental Panel on Climate Change (IPCC) for future emissions of the greenhouse gases and the atmospheric stabilization framework, a model used in support of the IPCC schemes³). Over the next century, warming would be reduced by about 1 °C, or 25 per cent of the expected warming post-1990. This reduced warming is similar to the reduction in warming predicted to result from stabilizing CO_2 emissions at 1990 levels; the reductions in warming from stabilizing methane concentrations or CO_2 emissions would be virtually identical until the year 2050.

Options for reducing methane

Opportunities for reducing methane emissions from each of the main anthropogenic sources have been identified and reviewed at IPCC meetings^{4,9,10}. It seems to be technically feasible to reduce methane emissions by around 120

Tg per year. Over the next 10 years, about one-third to one-half of these reductions would be needed to achieve the necessary 40–60 Tg reduction to stabilize atmospheric methane concentrations. The options for stabilizing methane emissions and realizing a range of benefits are as follows.

Landfills. Landfills currently account for about 30–70 Tg of methane, primarily from developed countries. Recovery systems can reduce emissions by 30–60 per cent in existing landfills by collecting the medium Btu gas resulting from the anaerobic decay of wastes⁴. The recovered gas may be burned directly in nearby industrial boilers, burned in internal combustion engines or turbines to generate electricity or, at a minimum, flared. New landfills may require liners to protect ground water from landfill leachate, allowing up to a 90 per cent reduction of methane emissions. Commercial operations in the United States, western Europe and other regions show that these systems can be profitable. Furthermore, the US government is proposing regulations to reduce emissions of air contaminants (primarily nonmethane organic compounds and air toxics) from landfills. As a side benefit, methane emissions from US landfills could be reduced by about 40–80 per cent¹¹.

Coal mining. Coal mining accounts for 30–50 Tg of methane, with most of the gas emitted from relatively few mines¹³. Degassing using vertical wells before and during the mining operation would reduce emissions of methane trapped in gassy underground mines by more than 50 per cent while reducing the costs of necessary mine ventilation^{4,13}. A portion of the recovered gas can be high Btu gas and fed directly into a pipeline or used to generate electricity. Gas of medium Btu quality recovered from gob wells over the mining operation could be used to generate electricity. Profitable examples of this technology exist in the United States, United Kingdom, Poland, Germany, China and Australia. Emissions may be reduced further by using mine ventilation air containing less than 1 per cent methane as combustion air in gas-fired turbines.

Oil and natural gas systems. Oil and natural gas systems account for about 25–45 Tg methane⁴. Most of the emissions probably occur in the developing countries, Eastern European countries

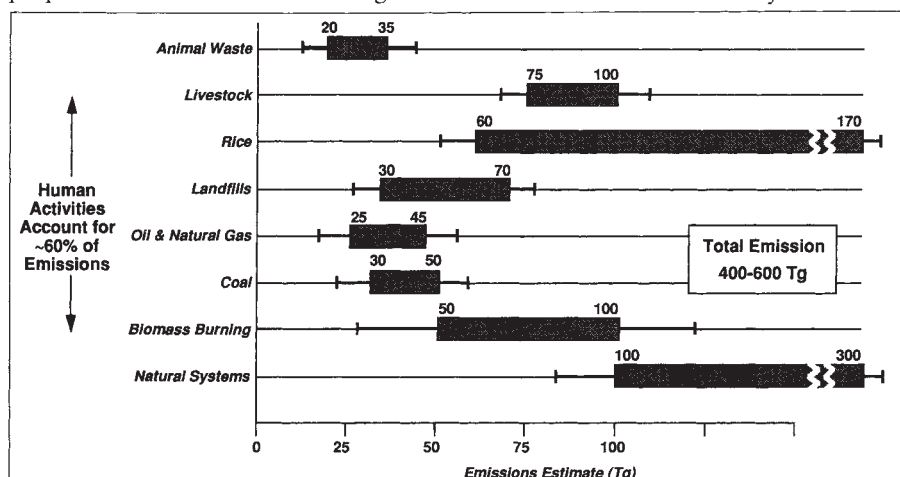


FIG. 1 Sources of methane. Hatched boxes, range of estimates; bold lines, additional uncertainty.

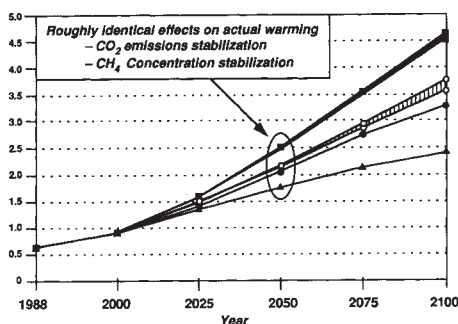


FIG. 2 Benefits of methane stabilization where methane emissions are capped at 540 Tg per year compared to capping CO₂ emissions at 1990 levels (with CO₂ concentrations rising to 500 p.p.m. by 2100). Ordinate, actual temperature increase, °C. Hatched portion, uncertainty range due to NO_x. Squares, IPCC-BAU; open circles, methane stabilization; closed circles, CO₂ capped at 1990 levels; triangles, methane and CO₂. Assumes 3° equilibrium warming.

and the Soviet Union. Improved handling of casing gas during oil production will reduce venting and flaring of natural gas. Improved technologies such as low-bleed pneumatic instruments, high-integrity piping and new pipeline-restoration techniques can also reduce methane emissions in a cost-effective manner⁴. In addition, emissions from gas transmission and distribution in the Soviet Union (where leakage is estimated to be about 6 per cent of throughput) and Eastern Europe could be reduced substantially.

Ruminants. Ruminants are the second largest anthropogenic methane source in the world, emitting 75–100 Tg (ref. 14), representing a lost opportunity to transform more carbon into useful products, such as meat or milk, during the natural fermentation of feed. In developed countries, specific feeds have been identified that may reduce methane emissions while enhancing productivity of cattle. In the United States, administration of bovine somatotropin, provided it receives regulatory approval, could increase milk production while reducing methane emissions by about 10 per cent. In other regions of the world, strategic supplementation programmes have increased milk production and reproduction efficiency, reducing methane emissions as a side benefit by perhaps as much as 60 per cent per gallon of milk¹⁵.

Animal wastes and wastewater. Animal wastes may contribute about 20–35 Tg (ref. 16) of methane and wastewater treatment an additional 25 Tg (ref. 4). Methane recovery systems can profitably capture 50 to 90 per cent of the methane emitted by anaerobic waste-management lagoons used mainly at dairy and swine operations. Methane recovered from lagoons may be used directly or to generate electricity for on-site demands or for selling back to an electrical grid¹⁷.

Rice cultivation. Rice cultivation may be the largest anthropogenic methane source. In the long term, methane emissions could probably be reduced by 10–30 per cent by an integrated management approach to irrigation, fertilizer application and cultivar selection¹⁸.

Biomass burning. Methane emissions resulting from biomass burning to clear new lands and after cropping are estimated at 50–100 Tg. These levels could be reduced by fire-management programmes and by encouraging the use of alternative agricultural practices^{4,19}.

How to make this work

These options for reducing methane emissions, while offering substantial potential for reductions, have not been implemented on a wide scale because of financial, informational and political barriers. In the United States, for example, although it can be profitable to recover and use methane that would have been emitted from coal mines (because of the value of the methane and reduced ventilation costs), institutional questions of ownership and constraints on receiving fair prices for gas or electricity may block implementation. Similarly, strategic supplementation of livestock can substantially increase productivity and create a local market for supplementation crops, but lack of capital and infrastructure may block implementation.

One component of an approach to reduce global methane emissions would be to identify existing barriers to the implementation of methane-reduction technologies and to establish national policies to overcome these barriers. For example, the People's Republic of China is considering the merits of changing the coal quota at state mines to allow the substitution of methane for coal, based on the BTUs produced. This will provide a strong incentive to recover methane.

A second component to global methane stabilization is identification of cost-effective projects for international lending institutions. Many methane-recovery projects during coal mining and

many livestock strategic supplementation projects are hampered only by lack of funds.

We suggest that the best approach would be identification of best-management practices for different methane sources and for different social and economic situations. For example, multinutrient molasses urea blocks can be recommended for livestock management in many developing countries; methane recovery and use can be recommended for landfills receiving large quantities of decomposable wastes as well as for coal mines, where the methane content of the coal and surrounding strata is high. These practices would vary widely from region to region, but would allow individual countries to adopt practices best suited to their needs and capabilities, to receive the value of the otherwise wasted methane, and would alleviate problems of monitoring and enforcement.

In conclusion, controlling methane emissions seems to be a feasible goal using current technology, and should be relatively cost-effective. As part of an approach to examine all greenhouse gases, coordinated international efforts are necessary to examine all greenhouse gases, coordinated international efforts are necessary to examine existing constraints that may hamper the implementation of cost-effective options for reducing methane emissions. Reductions in methane emissions appear to be very effective in mitigating radiative forcing in the atmosphere and eventual global warming. Reductions are also likely to moderate large-scale increases in tropospheric ozone and to counter the suppression of OH. Thus, as well as evaluating radiative effects of methane controls, it is important to consider the other roles of methane in the atmosphere. □

Kathleen B. Hogan and John S. Hoffman are in the Office of Air and Radiation, US Environmental Protection Agency, Washington, DC 20460; Anne M. Thompson is at the NASA/Goddard Space Center, Greenbelt, Maryland 20771, USA.

- Blake, D. R. & Rowland, F. S. *Science* **239**, 1129–1131 (1988).
- Shine, K. P., Derwent, R. G., Weubbles, D. J. & Morcrette, J.-J. in *Climate Change: The IPCC Scientific Assessment* (eds. Houghton, J. T., Jenkins, G. J. & Ephraums, J. J.) 41–68 (Cambridge University Press, 1990).
- Tirpak, D. & Lashof, D. (eds) *Policy Options for Stabilizing Global Climate* (EPA, Washington, DC, 1990).
- IPCC *Methane Emissions and Opportunities for Control* (September 1990).
- Crutzen, P. J. in *The Geophysiology of Amazonia* (ed. Dickinson, R. E.) 107–130 (Wiley, New York, 1987).
- Thompson, A. M., Stewart, R. W., Owens, M. A. & Herwehe, J. A. *Atmos. Envir.* **23**, 519–532 (1989).
- Watson, R. T., Rodhe, H., Oeschger, H. & Siegenthaler, U. in *Climate Change: The IPCC Scientific Assessment* (eds. Houghton, J. T., Jenkins, G. J. & Ephraums, J. J.) 1–40 (Cambridge University Press, 1990).
- Dak Sze, N. *Science* **255**, 673–675 (1992).
- Harris, R., Thompson, A. M., Gibbs, M. & Hogan, K. *Proc. NATO Adv. Res. Workshop Global Cycle of Atmos. Methane* (in the press).

- Greenhouse Gas Emissions from Agricultural Systems* (IPCC, EPA, 1990).
- Energy and Industry Subgroup Report* (IPCC, EPA, 1990).
- Fed. Reg.* **56**, No. 104, 24468–24528 (1991).
- Boyer, C. M., Kelafant, J. R., Kuuskraa, V. A., Manger, K. C. & Kruger, D. *Methane Emissions from Coal Mining* (EPA, Washington, DC, 1990).
- Crutzen, P. J., Aselmann, I. & Seiler, W. *Tellus* **38B**, 271–284 (1986).
- Leng, R. A. *Improving Ruminant Production and Reducing Methane Emissions from Ruminants by Strategic Supplementation* (EPA, Washington, DC, 1991).
- Casada, M. E., Safley, L. M. Jr, & Roos, K. *Global Methane Emissions from Livestock and Poultry Manure* (EPA, Washington, DC, 1991).
- Roos, K. in *Proc. Am. Soc. Agric. Engng* (in the press).
- Hogan, K. & Braatz, B. (eds.) *Sustainable Rice Productivity and Methane Reduction: Research Plan* (EPA, Washington, DC, 1990).
- US Agency for International Development *Greenhouse Gas Emissions and the Developing Countries Report to Congress* (July 1990).

Membrane proteins and minimalism

Nature's Boston conference last week* showed that many intracellular signalling pathways have succumbed to the attack of genetics and biochemistry, while the first mice without prions promise the solution of an outstanding mystery.

THE growing understanding of the molecular mechanisms of biological signal transduction has been accompanied by the development of an increasingly sophisticated iconography for representing the essential features of transmembrane receptors, or the chemical logic of their downstream signalling components, with the elegance and parsimony that come only with conceptual mastery. Some such reflection inspired Joseph Goldstein (University of Texas) to open this year's conference in Boston with a Mondrian.

But, inevitably, progress quickly overwhelms clean outlines by biochemical detail, and then there emerge new considerations, such as the role of prenylation (which is intelligible to biochemists and cell biologists) and the phenomenon of infectious prions (which, so far, is not).

Much of the most spectacular reductionist success has been in the functional dissection of membrane proteins and the definition of the biochemical pathways whereby receptors signal to the interior. The most elegant example remains the work of Joseph Goldstein and Michael Brown (University of Texas) on the low-density lipoprotein (LDL) receptor.

This has now branched in at least two directions. One has led Brown to the analysis of the LDL-related receptor — a large, multifarious and possibly ancestral membrane protein that uses its repeated sequences both to bind different apoproteins and to mop up serine protease complexes from the bloodstream. The other route, via the mevalonate metabolic pathway, has in just six months (according to Goldstein) opened a new and unsuspected field for the exploration of possible specific drugs; perhaps as many as half of all cytoplasmic proteins (including *ras*, the first example) have lipid tails which profoundly influence their localization and function.

Interference with ligand binding at the cell surface remains the focus of vigorous investigation, illustrated by attempts to arrest destructive neutrophils with anti-ICAM-1 (Timothy Springer, Harvard) or to attack retroviral leukaemia with antibodies against components of the IL-2 receptor (Thomas Waldman, National Institutes of Health). But hope of new specific drug targets increasingly centres on the cytoplasmic domains of receptors, and on the downstream elements with which they interact.

It is now possible to identify small regions in the cytoplasmic tails of recep-

tors (or receptor complexes) mediating interactions with downstream signalling components in the β -adrenergic receptor (Robert Lefkowitz, Duke University), the interleukin-2 receptor (Tadatsugu Taniguchi, Osaka University) and the new family represented by the erythropoietin receptor (Harvey Lodish, Whitehead Institute). In some cases, domains linked with different intracellular signalling pathways can be distinguished, as in the antigen receptor of T cells (Richard Klausner, National Institutes of Health) and in the platelet-derived growth factor in which Lewis Williams (University of California) has been able to show specific binding of PI3 kinase, phospholipase C and GAP via SH2 motifs to specific phosphorylated tyrosines of the receptor.

Yet in spite of the emerging pattern of the first intracellular steps in signal transduction, many questions remain about possible specific interactions between downstream signalling elements, or branching and cross-talk between pathways. Thus many signal transduction pathways lead to a change of cyclic AMP concentration, which may be achieved by activation of any of a number of different adenylyl cyclases (Alfred Gilman, University of Texas). These have been generally thought to be activated by the α -subunit of the heterotrimeric G protein; but experiments by Gilman *in vitro* and by Henry Bourne (University of California) in transfected cells seem to show that adenylyl cyclase II can be activated by the $\beta\gamma$ -subunit of an inhibitory G protein in the presence of the α -subunit of an excitatory one.

Only rarely can biochemical experiments demonstrate unequivocally that a particular association or reaction is physiologically relevant, especially at the level of the whole animal, hence the telling power of genetics, which has dramatically demonstrated in human neoplastic disease the physiological relevance of p53 (Bert Vogelstein, Johns Hopkins University) and of the GAP protein NF1 (Frank McCormick, Cetus Corporation). Genetics has also led to the identification of the cystic fibrosis transmembrane conductance regulator (CFTR) and the remarkable pleiotropic interaction of the c-kit tyrosine kinase receptor and its variously named ligand, mutations in either of which lead to anaemia, sterility and albinism (Peter Besmer, Sloan-Kettering).

Yet in human mutations, the relationship between genotype and phenotype may

be elusive. Despite the elegant functional dissection of the CFTR by mutagenesis *in vitro*, the origin of variations in the severity of the phenotype is not known (Michael Welsh, University of Iowa), nor is the biochemical pathway whereby NF1 inhibits the growth of Schwann cells and melanocytes.

In defining biochemical pathways, the genetics of fruitflies, nematodes and yeast has been dominant. Gerry Rubin (University of California) showed how the fruitfly ommatidial mutation *sevenless* has led to the definition of a *ras*-dependent pathway in normal development as had the genetics of nematode vulval morphogenesis (Robert Horvitz, MIT). Polarization mutants in yeast seem likely to reveal cytoskeletal regulatory mechanisms that may bear on mammalian morphogenesis (Ira Herskowitz, University of California).

Higher neural functions made their appearance at Boston as olfaction and as what are now known as prion diseases. The identification of an astonishingly large number of receptors for odorants raises the question of how the signal from a specific ligand is interpreted as a distinct sensation — a problem in neural coding, not transmembrane transduction (Richard Axel, Columbia).

Prion diseases can be either genetic — a mutation in the gene encoding the prion protein — or infectious, when a mutant prion protein is injected into the brain of a normal animal. In either case, the result is neural degeneration by a mechanism not understood. But transgenic mice containing mutant prion genes develop degenerative disease that can be transmitted with the protein to other animals; so the phenomena are incontrovertible (Stanley Prusiner, University of California).

Of the theories to explain them, one of the most prominent is that mutant prion proteins adopt an abnormal conformation transmitted when they interact with the normal wild-type version. So an animal lacking wild-type prion protein should be immune from prion disease. Charles Weissmann (Zurich) announced in Boston that he is now able to test this by means of homozygous "knockout" mice in which the prion gene is not expressed and which so far seem phenotypically normal. This time next year, it should be known whether they will succumb to infection with mutant prions.

Miranda Robertson

*"Membrane proteins and signal transduction", 14 & 15 November 1991, Boston.

New variations on the theme

C. J. Farr and P. N. Goodfellow

THE latest results from Alec Jeffreys' laboratory (page 204 of this issue¹) will attract the close attention of patent officers and the legal profession, as well as scientists. The paper reports a new DNA-fingerprinting technique which extends existing methods by assaying sequence variation within minisatellite arrays.

Minisatellites are hypervariable regions of the genome which exhibit polymorphism due to a variable number of tandem repeats of a short sequence. In 1985, Jeffreys and co-workers²⁻⁴ demonstrated that a DNA 'polycore' probe based on a myoglobin gene minisatellite can simultaneously detect, by hybridization to human genomic DNA, a large number of dispersed hypervariable loci containing tandem repeats of similar sequence. The complex restriction-fragment pattern that results can produce a DNA fingerprint specific to an individual (the only exception being monozygous twins), and the applications have been legion.

To improve the sensitivity of multilocus DNA fingerprint analyses, the next step was amplification of human minisatellites by the polymerase chain reaction (PCR). Perhaps surprisingly, in view of their highly repetitive nature, it was found that *Taq* polymerase could amplify entire minisatellites whilst preserving the allelic specificity of the number of repeat units⁵. Moreover, several minisatellites could be co-amplified to generate reproducible PCR-derived DNA fingerprints, increasing sensitivity by several orders of magnitude, and high-fidelity amplification of hypervariable loci in one or a few cells became attainable.

MVR mapping

Last year PCR was used to analyse allelic variation within a minisatellite locus⁶, until then an unexplored source of human DNA polymorphism. This approach, called minisatellite variant repeat (MVR) mapping, dramatically increases the number of different alleles of any minisatellite locus that can be distinguished in a population. This method has been applied to the hypervariable locus *DIS8* (probe MS32) at region 42-43 on the long arm of chromosome 1. Previous DNA sequence analyses of cloned MS32 had shown that about half the repeat units share an A→G transition which creates an *HaeIII* restriction site readily amenable to MVR mapping. Interspersion patterns of *HaeIII*⁺ and *HaeIII*⁻ repeat units (designated a-type and t-type) were assayed by PCR ampli-

fication of an entire allele followed by end-labelling and partial digestion with either *HaeIII* or *HinfI*, agarose gel electrophoresis and autoradiography.

The partial digest pattern generated by *HinfI*, which cuts once within each repeat unit, produces a continuous ladder of labelled fragments from which the number of minisatellites can be determined. Comparison of *HaeIII* and *HinfI* partial digests enables each repeat unit to be scored as to whether or not it is cleaved by *HaeIII*. High levels of allelic variation were revealed, particularly near the beginning of the *DIS8* array. MVR mapping provides an unambiguous binary code for any allele, but is technically demanding (for mere mortals at least), limited to alleles small enough to amplify by PCR, and cannot be used for diploid mapping.

Jeffreys *et al.*¹ have now come up with a much simpler system, MVR-PCR. It involves the use of two MVR-specific oligonucleotides which differ at their 3' ends, allowing allele-specific priming of either a-type or t-type repeat units. PCR using one or other of these primers, together with a primer from a fixed site in the minisatellite-flanking DNA, generates two complementary sets of products from which the MVR map can be deduced. To prevent the oligonucleotides priming internally within PCR products, MVR detection and subsequent amplification are uncoupled by providing each MVR-specific primer with a 20-nucleotide 'tag' at the 5' end. Primers complementary to the tag are used for amplification in subsequent cycles.

MVR-PCR can be undertaken on total genomic DNA, where a superimposed profile of both alleles produces a diploid code, and the DNA profiles are generated as extraordinarily variable digital sequences resembling bar codes suitable for computer analysis. The data are highly objective, each experiment being internally controlled. In addition the method is reliable and sensitive, and can be applied to degraded DNA, making it particularly well suited to forensic work. Many of the problems associated with conventional DNA fingerprinting, such as the need for side-by-side comparison of DNA samples of the same gel, are obviated.

But what of the limitations? One is that low-cycle PCR prevents direct visualization of the PCR products in ethidium bromide-stained gels. More serious is that only a few minisatellites have so far been identified that are suitable for analysis by this approach —

that is, minisatellites that show internal variation, but do not display variant length repeats (abnormal length repeats would throw the MVR diploid coding ladders out of register and make them uninterpretable). Moreover, the extremely high *de novo* mutation rate at *DIS8* limits the effectiveness of the method in, for example, paternity testing and evolutionary studies, two areas for which, at first glance, it would appear to be ideal. Also, in establishing family relationships, the possibility of error arises from the need to identify heterozygous 'null' positions from interpretation of band intensities (dosage is important where one allele contains a sequence variant which fails to amplify with either primer).

Allelic variability

The resolving power of MVR-PCR reveals that there can be very high levels of allelic variability at human minisatellites and provides the first evidence that recombination is involved in the generation of ultravariability. Several lines of evidence^{2,7-9} have suggested that minisatellites may be recombination hotspots involved in chromosome pairing or the initiation of meiotic recombination, or both. But the failure to detect unequal crossing-over between homologous chromosomes^{6,10,11} seemed to contradict that notion. The finding¹ of two apparent recombinants at *DIS8* in only 572 meioses tested will re-open debate about the involvement of minisatellites in meiosis and recombination.

In British legal history, the name of Jeffreys has been indelibly associated with the judge at the Bloody Assizes of 1685, and with arbitrary and brutal 'justice'. His latter-day (presumably unrelated) namesake is bringing scientific objectivity to the law courts. In part as eponymous tribute, in part as a gesture against the rising tide of acronyms, we propose that MVR-PCR should be known as the Jeffreys method. □

C. J. Farr and P. N. Goodfellow are in the ICRF Laboratories, Lincoln's Inn Fields, London WC2A 3PX, UK.

1. Jeffreys, A. J., MacLeod, A., Tamaki, K., Neil, D. L. & Monckton, D. G. *Nature* **354**, 204-209 (1991).
2. Jeffreys, A. J., Wilson, V. & Thein, S. L. *Nature* **314**, 67-73 (1985).
3. Jeffreys, A. J., Wilson, V. & Thein, S. L. *Nature* **316**, 76-79 (1985).
4. Jeffreys, A. J., Brookfield, J. F. Y. & Semeonoff, R. *Nature* **317**, 818-819 (1985).
5. Jeffreys, A. J., Wilson, V., Neumann, R. & Keyte, J. *Nucleic Acids Res.* **16**, 10953-10971 (1988).
6. Jeffreys, A. J., Neumann, R. & Wilson, V. *Cell* **60**, 473-485 (1990).
7. Steinmetz, M., Stephan, D. & Lindahl, K. F. *Cell* **44**, 895-904 (1986).
8. Royle, N., Clarkson, R. E., Wong, Z. & Jeffreys, A. J. *Genomics* **3**, 352-360 (1988).
9. Chandley, A. C. & Mitchell, A. R. *Cytogenet. Cell. Genet.* **48**, 152-155 (1988).
10. Wolff, R. K., Nakamura, Y. & White, R. *Genomics* **3**, 347-351 (1988).
11. Wolff, R. K., Plaetke, R., Jeffreys, A. J. & White, R. *Genomics* **5**, 382-384 (1989).

Justifying the means

Ian Stewart

WHO was the leading mathematician of all time? To many mathematicians, the top candidate is Carl Friedrich Gauss, a bricklayer's son born in 1777 in Brunswick. So wide-ranging were Gauss's ideas, and so high were his standards of publication, that later mathematicians repeatedly found that Gauss had anticipated their life's work but kept it to himself. Examples are Bolyai and Lobachevskii's discovery of non-euclidean geometry, Cauchy's work in complex analysis, and Jacobi's theory of elliptic integrals. It would not be a great surprise if an obscure manuscript by Gauss on the Mandelbrot set were to turn up.

Although no such work is known, Gauss had some closely related ideas, which are now being revived as the Mandelbrot set focuses attention on what is now called complex dynamics. They are interpreted from the modern viewpoint, and generalized, in a recent paper by Shaun Bullett (*Topology* 30, 171–190; 1991).

The arithmetic mean of two numbers is half their sum; the geometric mean is the square root of their product. Gauss invented a mixture of the two, which he called the arithmetico-geometric mean (a.g.m. for short). Start with a pair of numbers (a, b) . Calculate their arithmetic and geometric means, to get a new pair $(a, b) = (\frac{1}{2}(a+b), \sqrt{ab})$; repeat this process indefinitely. The two numbers converge to a common limit, the a.g.m. of the original numbers. This is a deceptively simple procedure: it conceals some remarkably deep mathematics.

Gauss kept a mathematical diary, and the entry for 30 May 1799 records that the a.g.m. of 1 and $\sqrt{2}$ is 2π divided by the total length of a figure-of-eight shaped curve known as a lemniscate, whose formula given in polar coordinates is $r^2 = \cos 2\theta$. He discovered this by evaluating the length to 11 decimal places. He correctly predicted that "the demonstration of this fact will open up an entirely new field of analysis". That field occupied the attention of many nineteenth-century mathematicians: it is the theory of elliptic functions, and it too is experiencing something of a revival, for various reasons.

Gauss had a great interest in complex numbers $x+iy$ where $i=\sqrt{-1}$. The above discussion of the a.g.m. assumes the initial numbers are real and positive, because square roots are involved in the

geometric mean, and negative real numbers do not have a real square root. All complex numbers possess square roots, so the concept of a.g.m. extends naturally to arbitrary pairs of complex numbers. However, any complex number other than 0 has two distinct square roots, so the calculation of the a.g.m. involves an arbitrary choice at each step. Gauss wondered what the set of all possible values of the a.g.m., obtained by making all possible sequences of choices of square root, would look like.

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Orbits of the a.g.m. correspondence reveal an intricate packing of Siegel disks (white regions).

The modern approach, adopted by Bullett, is to view his procedure as a 'quadratic correspondence', a two-valued map $z \rightarrow w$. Suppose a and b are multiplied by some real constant. Then so are the arithmetic and geometric means. In other words, we can always scale the value of b to equal 1. Then the pair $(z, 1)$ leads to a new pair $(w, 1)$, given by the quadratic equation $4w^2z - z - (z+1)^2 = 0$. Call this the a.g.m. correspondence. It is an example of a more general concept, that of a quadratic correspondence. Conventional discrete dynamics studies the effect of iterating a single-valued mapping; here a two-valued mapping must be iterated, but the situation is closely analogous. The famous Mandelbrot set, and its relatives, the Julia sets, arise from conventional dynamics in the complex plane. Gauss was studying a two-valued analogue.

Dynamicists study orbits — the sets traced out by a given initial point under iteration of a mapping. The simplest orbits are fixed points, which map to themselves; the next simplest are periodic points, which repeat a cycle of

values; and at the other extreme are chaotic orbits that display no obvious pattern whatsoever. The fixed points for the a.g.m. correspondence are 1 and ∞ . The point 1 is attractive: nearby points move towards it when the mapping is applied. The point ∞ is repulsive.

But the dynamics is far more intricate than just a steady flow away from ∞ towards 1, as Bullett's computer plot (see figure) of a restricted set of orbits illustrates. The blank regions in this picture are called Siegel disks. At the centre of each Siegel disk lives a periodic point; and these points are surrounded by closed curves, along which points move much as if the curves were rotating through various angles at each application of the mapping. The fractal nature of the picture is evident, and similar Siegel disks can be seen in *The Beauty of Fractals* by H.-O. Peitgen and P. Richter (Springer, New York, 1988), a book that is largely about Mandelbrot and Julia sets.

The connection with the lemniscate led Gauss in a rather different direction (though one might speculate on what would have happened if there had been a workstation on his desk). While trying to calculate the length of the curve, he invented what are now called theta functions, subsequently developed extensively by Jacobi. They are defined by certain infinite series, for example

$$p(t) = \sum \exp(n^2 \pi i t)$$

$$q(t) = \sum (-1)^n \exp(n^2 \pi i t)$$

summed over all positive and negative integers n . They satisfy some remarkable identities; in particular the arithmetic mean of $p(t)^2$ and $q(t)^2$ is $p(2t)^2$, and their geometric mean is $q(2t)^2$. In short, if we start the procedure for computing the a.g.m. not with (a, b) but with $(p(t)^2, q(t)^2)$, then the next step just replaces t by $2t$. The map $t \rightarrow 2t$ thus 'induces' the a.g.m. correspondence, but is far simpler to think about.

Using his version of these ideas, Gauss came very close indeed to solving his question about the set of values produced by making arbitrary choices of square roots in the a.g.m. calculation. Suppose the starting values are (a, b) . Let λ and μ be the 'simplest' values of a.g.m. (a, b) and a.g.m. $(a+b, a-b)$. Then the other possible values μ' of a.g.m. (a, b) are given by $1/\mu' = d/\mu + ic/\lambda$, where c and d are integers with no common factor, c is a multiple of 4, and $d-1$ is a multiple of 4. This may seem a complicated answer, but it is direct and explicit once the theta functions have been introduced. In a sense, their role is to 'linearize' the a.g.m. correspondence by a cunning change of variable. The conditions on c and d also reveal deep

Shaun Bullett

connections with number theory.

In a recent preprint ("Critically finite correspondences and subgroups of the modular group", Queen Mary and Westfield College, 1991), Bullett obtains a remarkable classification of a more general class of correspondences with strongly constrained dynamics. Among them is a finite set of quadratic correspondences having curious connections

GALAXIES

Inner secrets of ellipticals

James Binney

FOR many years elliptical galaxies were thought to be the most 'relaxed', that is bland, featureless and boring, of galaxies. But during the 1980s it emerged that far from being boring, many ellipticals are downright zany. Two papers in this issue describe computer simulations which explain some of the more intriguing complexities of elliptical galaxies.

Foremost among these complexities is the observation that the main body of a galaxy in many cases rotates about a completely different axis from its nucleus — the nucleus of NGC5322 actually rotates in the opposite direction. Also, deep inside these galaxies one can often just detect gas on nearly circular orbits like the gas that forms the disk of the Milky Way. But in ellipticals, the spin axis of the gas often varies abruptly from one radius to another, so that the overall effect is of misaligned annuli rather than a coherent disk. How do these systems come to be so disorganized that they can harbour counter-rotating cores or annuli of gas that rotate about mutually perpendicular axes?

L. Hernquist and J. E. Barnes (page 210) have been using 'N-body' simulations to model what happens when two spiral galaxies merge, the process that most astronomers now think is responsible for forming elliptical galaxies. N-body simulations of mergers have been a minor industry for 15 years and it has been known for some time that when two numerical disk galaxies merge, the *N* 'stars' soon arrange themselves into a fair representation of the main body of an elliptical galaxy. What makes Hernquist and Barnes's simulations unusual is that they include particles which interact with one another in a way that represents the dynamics of gas. And lo, these 'gas' particles arrange themselves into a central disk and a surrounding annulus which rotates in the opposite sense to the disk. Hernquist and Barnes suggest that this is how the puzzling counter-rotating systems in NGC5322 and other galaxies were formed.

Could it be that the sharp distinction in so many galaxies between the kinema-

tics of the nucleus and of the galaxy at large is a consequence of the very different responses of stars and gas to the violently fluctuating gravitational field that accompanies a merger? In particular, are the stars in the nucleus younger than those in the surrounding galaxy because they formed after the last merger? In principle, the age of a stellar population can be determined from the spectrum of the light it emits, but in practice the determination is difficult even for the nearest galaxies and so far there is no definite answer either way. But the prospect of being able to date the last substantial merger of a galaxy is an exciting one.

Ian Stewart is in the Mathematics Institute, Warwick University, Coventry CV4 7AL, UK.

It is thought that the nuclei of many giant galaxies contain massive black holes which at some time powered violent nuclear activity of the sort that gives rise to quasars, radio galaxies, Seyfert galaxies and the rest. T. Ebisuzaki, J. Makino and S. K. Okumura (page 212) have been using a purpose-built parallel computer to study how the presence of black holes in the nuclei of the progenitor galaxies modifies the structure of the merger remnant.

When a massive object moves through a field of stars, its motion is opposed by dynamical friction, a gravitational drag that arises from the tendency of the massive object to scatter passing stars into orbits of higher energy. Consequently, during a merger massive black holes from the progenitor galaxies would sink rapidly towards the centre of the remnant, and then go into orbit about one another. Dynamical friction would cause the orbit of this binary hole to lose energy and shrink. In previous analytical work it was assumed that as the orbit shrinks it becomes more and more circular. Ebisuzaki *et al.* find that it actually becomes more and more eccentric and thus the closest approach to the centre on each orbit rapidly decreases. Eventually, gravitational waves excited by successive passages past the centre carry off more energy than is lost to dynamical friction, and the holes quickly merge.

There are two important aspects to

this description. First, it is important that the holes are predicted to merge quite rapidly as this ensures that if the remnant is involved in a further merger, neither of the merging galaxies is likely to contain a binary black hole: if one of them did, a triple black hole would be formed and this would soon spontaneously 'ionize' — eject one of the triple — in the gravitational analogue of the Auger effect of atomic physics. If a triple hole were to ionize in this way, both the ejected hole and the surviving binary would almost certainly be ejected at great speed from the new merger remnant, never to return. By ensuring an early merger of the holes, Ebisuzaki *et al.* make it more plausible that even in the popular cold-dark-matter picture of galaxy formation, in which large structures are built up by repeatedly merging smaller objects, most giant galaxies contain a massive black hole.

The second important aspect of Ebisuzaki *et al.*'s picture is its ability to account for a correlation between the total luminosities of elliptical galaxies and their core radii. In crude terms, the core radius is the largest radius inside which the luminosity density is nearly constant. When two galaxies that have no central black holes merge, the remnant's core radius is, if anything, a bit smaller than the core radii of the progenitor galaxies. So if large galaxies were built up by repeatedly merging small ones, the core radii of galaxies would not be positively correlated with their luminosities, as is observed to be the case. The simulations by Ebisuzaki *et al.* show that when the progenitor galaxies contain black holes, the energy released by the short-lived binary hole as its orbit decays by dynamical friction, causes the remnant's core radius to increase by just the right amount to account for the observed correlation with luminosity.

The papers by Hernquist and Barnes and by Ebisuzaki *et al.* are concerned with dynamics on the smallest galactic scales. But gravity has no natural scale, and the structure of typical galactic potentials is such that the cooling characteristics of gas, which do define a characteristic temperature, scarcely pick out a characteristic length scale either. So it is tempting to rescale these results by a factor of 10–100 and apply them to the merger of the dark halos that encompass galaxies.

If, as is now the fashion, one believes that the Universe has the critical density for gravitational closure, it is hard to escape the conclusion that the dark halo of a galaxy like our own stretches out to at least 700 kiloparsecs — for comparison, the distance of the Sun from the Galactic Centre is about 8.5 kiloparsecs. So, viewed on the vast scale of a dark halo, a visible galaxy is no more than a

pip, a nucleus. Do we infer from the work of Hernquist and Barnes that it will often be kinematically decoupled from its surrounding halo? More than half of all galactic disks develop 'warps' at radii of the order of 10 kiloparsecs and a popular explanation of the phenomenon (A. Toomre, *IAU Symp.* **100**, 177–186; 1983) is precisely that the spin axes of galaxies are misaligned with those of their halos. Moreover, rings of gas are sometimes found around disk galaxies, the spin axes of the ring and the disk being roughly perpendicular just as in Hernquist and Barnes's simulation.

Finally, a potential problem with the

cold-dark-matter model is that large halos that have grown by repeated accretion would be expected to have very small core radii, if any. But such coreless halos are probably incompatible with the observed rotation curves of many disk galaxies. When two halos merge, do the visible galaxies act like the black holes in the calculations by Ebisuzaki *et al.* and increase the core radius of the halo at each stage in the merging process? □

James Binney is in the Sub-department of Theoretical Physics, Oxford University, 1 Keble Road, Oxford OX1 3NP, UK.

NEUROSCIENCE

Vehicles of inactivation

Solomon H. Snyder

FOUR reports in *Nature*¹ and *Science*^{2–4}, with another to come in *Proceedings of the National Academy of Sciences*⁵, together announce the cloning of two of the major transporter molecules for neurotransmitters. These developments are the latest fruit of the explosion in molecular studies of transmitter transport; they will, one hopes, add further impetus to the drive for psychotherapeutic innovation.

After their synaptic release, biogenic amine and amino-acid neurotransmitters are inactivated by transport back into the nerve terminals that had released them. Therapeutic agents (such as tricyclic and related antidepressants) and drugs of abuse (such as cocaine) seem to act by blocking this transport for one or another biogenic amine. Molecular cloning, which has provided deep insight into postsynaptic receptors, has quite recently been applied to presynaptic neurotransmitter transporters or 'uptake receptors'. Now, essentially simultaneously, two groups report the cloning of serotonin transporters^{1,2}, and three the cloning of dopamine transporters^{3–5}.

As the first neurotransmitter to be

characterized, acetylcholine historically provided the model for transmitter disposition. Because acetylcholine is inactivated through enzymatic degradation by acetylcholinesterase, it was assumed for many years that synaptic inactivation for all transmitters would involve enzymatic breakdown. In the late 1950s and early 1960s, Julius Axelrod and colleagues overturned this idea, showing that transmitter re-uptake was the main inactivator for catecholamines⁶. Monitoring of radiolabelled transmitter accumulation by isolated nerve terminals (synaptosomes) extended study of the importance of uptake to other biogenic amines and amino-acid transmitters⁷. Re-uptake is now regarded as the universal mechanism of transmitter inactivation, enzymatic degradation applying only to acetylcholine and neuropeptides.

As with many postsynaptic receptors, the molecular cloning of presynaptic transporters resulted from isolation of the relevant transport protein, a task first accomplished by Kanner and associates, who purified the γ -aminobutyric acid (GABA) transporter⁸ and then cloned its complementary DNA (ref. 9).

GABA transport was expressed by transfecting the single cloned cDNA encoding a protein of relative molecular mass 67,000, which, like the glucose and other transporters, displays 12 transmembrane regions, but does not show sequence similarity to these other transporters. Amara and colleagues used expression cloning strategies to isolate and sequence cDNA for the noradrenaline transporter¹⁰. The 70 per cent similarity in amino-acid sequence of the two transporters encouraged several groups to adopt polymerase chain reaction (PCR) technology based on conserved elements of the GABA and noradrenaline transporters, resulting in the cloning of dopamine and serotonin transporters; a comparison of the four is shown in the table. The 60–70 per cent similarity in the amino-acid sequences of GABA and the three amine transporters portends the cloning in the near future of all known transmitter transporters.

The cloned transporters should help add to our knowledge of drug action; indeed, the papers on the dopamine^{3–5} and noradrenaline¹⁰ transporters give headline prominence to the transporters' sensitivity to cocaine. Among a series of cocaine derivatives, inhibition of dopamine uptake correlates best with behavioural reinforcing effects in animals¹¹. But the potencies of cocaine at noradrenaline, dopamine and serotonin transporters are quite similar, so that at doses ingested by humans all three uptake systems will be influenced. Among therapeutic agents, antidepressants are the most prominent drugs that act by blocking transmitter uptake, with resultant potentiation of noradrenaline and serotonin as a therapeutic mechanism. Inhibition of GABA and glycine uptake should provide sedative, anti-anxiety and anticonvulsant effects, but adequately potent and safe agents have not yet appeared. With hard work and ingenuity, and maybe a little luck, it may not be long before they do. □

Solomon H. Snyder is in the Departments of Neuroscience, Pharmacology and Molecular Sciences, and Psychiatry, The Johns Hopkins University School of Medicine, 725 N. Wolfe Street, Baltimore, Maryland 21205, USA.

Neurotransmitter transporters				
	GABA	Noradrenaline	Dopamine	Serotonin
Relative molecular mass	67,000	69,000	69,000	73,000 ¹ , 68,000 ²
Amino acids	599	617	619	653 ¹ , 607 ²
Messenger RNA (kilobases)	4.2	5.8, 3.6	3.7	3.1 ¹ , 3.7 ²
Glycosylation	3	3	4	2
PKA sequence	1	0	0	1

All four transporters display 12 transmembrane domains with both C and N termini being cytoplasmic. The absence in noradrenaline and dopamine transporters of canonical sequences for phosphorylation by cyclic-AMP-dependent protein kinase (PKA) does not rule out PKA phosphorylation, as these two transporters possess less stringent consensus sites for PKA phosphorylation. Differences in length of the two serotonin transporter proteins reflect isolated base differences in sequences from the two laboratories^{1,2} which shift the reading frame. Whether cloning errors are involved or whether there are real differences between the serotonin transporters of rat brain¹ and rat basophilic leukocytes² is not apparent.

1. Blakely, R. D. *et al. Nature* **354**, 66–70 (1991).
2. Hoffman, B. J., Mezey, E. & Brownstein, M. J. *Science* **254**, 579–580 (1991).
3. Shimada, S. *et al. Science* **254**, 576–578 (1991).
4. Kilty, J. E., Lorang, D. & Amara, S. G. *Science* **254**, 578–579 (1991).
5. Usdin, T., Mezey, E., Chen, C., Brownstein, M. J. & Hoffman, B. J. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
6. Axelrod, J. *Science* **173**, 598–606 (1971).
7. Snyder, S. H. *Biol. Psychiat.* **2**, 367–389 (1970).
8. Radian, R., Bendahan, A. & Kanner, B. I. *J. biol. Chem.* **261**, 15437–15441 (1986).
9. Guastella, J. *et al. Science* **249**, 1303–1306 (1990).
10. Pacholczyk, T., Blakely, R. D. & Amara, S. G. *Nature* **350**, 350–354 (1991).
11. Ritz, M. C., Lamb, R. J., Goldberg, S. R. & Kuhar, M. J. *Science* **237**, 1219–1223 (1987).

A deep look at African rifting

Paul Morgan

THE East African rift valley provides geologists with a modern example of the processes that have split continents over the eons and led to the growth of new oceans. Reports elsewhere in this issue^{1,2} from two phases of the Kenya Rift International Seismic Project (KRISP 85 and 90) suggest from the structure of the rift far below the surface that the plastic asthenosphere below the Earth's crust plays an active part in the rifting.

Continental rifts are a fundamental tectonic feature of the continents, the sites of future or failed continental separation and the birth of new oceans. They are easily identified by their surface morphology as elongated depressions bounded by normal faults in the upper crust. The common association of rifts with volcanism, high heat flow, anomalous crust and upper-mantle structure, and seismic activity provides compelling evidence that the processes of rifting are not confined to upper crustal levels, but are linked to dynamic processes in the usually rigid outer lithosphere and the underlying weak asthenosphere. Since agreement was reached, over half a century ago, that rifts are generally dominated by extensional processes, two fundamental questions have remained concerning the mode and mechanism of continental rifting. What is the primary mode of extensional strain during continental rifting (irrotational pure strain or simple strain, with rotation)³. And is the asthenosphere active or passive in the mechanism of continental rifting^{4,5}?

Two phases of a multinational research project have provided the first relatively detailed models of lithospheric structure beneath the Kenya rift, a prime example of the rifting process, based upon seismic waves recorded from man-made explosive and natural earthquake sources. The first phase of the project, KRISP 85, demonstrated the applicability of seismic studies to this problem by providing evidence of three-dimensional complexity in this grossly two-dimensional structure⁶, and refining earlier studies of thinning of the lithosphere beneath the rift zone⁷. The second phase, KRISP 90, was primarily designed to study variations in crustal thinning along and across the rift axis.

Verney Green *et al.*¹, reporting on page 199, use arrival times from KRISP 85, and from their own experiments for compressional seismic waves from distant earthquakes, to confirm and refine the earlier studies^{7,8} by mapping significant velocity anomalies in the lower crust and upper mantle beneath the Kenya rift. In the lower crust, their

results indicate a 12 per cent velocity range, with the highest velocities beneath the rift axis, but significant variation in both the width and magnitude of this high-velocity anomaly along strike of the rift. In the uppermost mantle, the authors find a 10–12 per cent decrease in velocity relative to the mantle in adjacent stable regions. The region of anomalous velocities is bounded by steep sides, and is traced from a depth of about 145 km up to 35–65 km directly beneath the rift.

The authors attribute the low-velocity anomaly to dome-like intrusions (diapirs) of hot asthenosphere in the lithosphere beneath the rift. Approximately half the anomaly is attributed to temperature contrasts between the lithosphere and asthenosphere, and half is thought to result from 3–6 per cent partial melting in the rising diapir generated by pressure-release melting. High velocities in the lower crust result from modification of the lower crust by magmas rising

from the asthenosphere diapir. Variations in the widths and magnitudes of the anomalies along the rift perhaps indicate multiple diapirs along the rift axis.

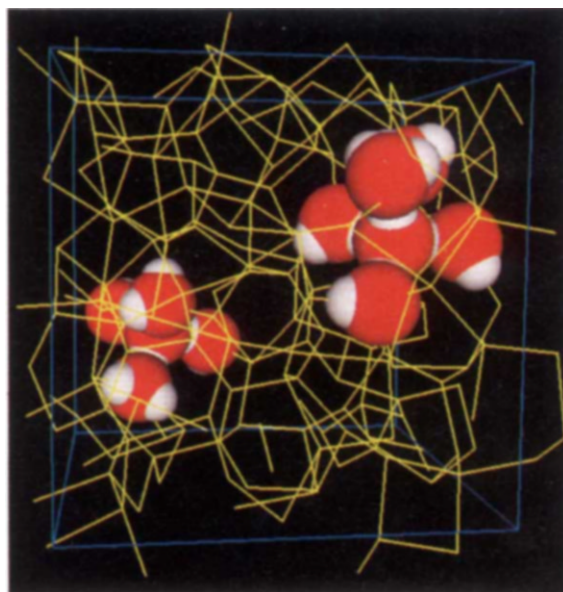
Preliminary results² from KRISP 90 reported on page 223 indicate further three-dimensional lithospheric structure associated with the Kenya rift. Using refracted and wide-angle reflected arrivals from explosive sources, the results indicate a thinning of the crust from 35 km in the south to 20 km in the north along the rift axis, and abrupt thickening of the crust by 6–10 km across the rift at the rift margins. A significant decrease in compressional wave velocities in the uppermost mantle, from 8.0–8.1 km s⁻¹ beneath the rift flanks to 7.5–7.7 km s⁻¹ beneath the rift axis, confirm other indications of anomalous upper mantle immediately beneath the rift axis⁶⁻⁸. In addition, the results indicate crustal thickening immediately west of the rift. This thickening is close to a major pre-rifting crustal boundary between the Archaean craton to the west, where the crust has been stable for 2.5 billion (10⁹) years, and the Pan-African orogenic belt which experienced major deformation about 550 million years ago.

Water goes with the flow

If hydrogen bonding is strong enough to cause the collapse of polymer gels and to bind acetic acid molecules into dimers in the gas phase, why does it not hold liquid water rigid? The example *par excellence* of a hydrogen-bonded system, water, consists of a continuous random network of tetrahedrally coordinated molecules not unlike that in vitreous silica, with bond energies much larger than the thermal energy at room temperature; yet it remains a liquid. On page 218, F. Sciortino *et al.* show that the fluidity of water can be ascribed to defects in the hydrogen-bonded network in which some molecules are coordinated to five neighbours, not the usual four.

To acquire the extra neighbour, one of the hydrogen bonds formed by the central molecule must be bifurcated, in that either one of its hydrogen atoms bonds to two oxygens or three of the neighbours' hydrogens are coordinated to its oxygen atom. The simulations of Sciortino *et al.* reveal the somewhat counterintuitive result that the higher density in the first coordination shell of five-coordinate molecules leads to increased molecular mobility. This is not, as might be thought, a consequence of weaker bonding, but of a 'flattening' of the potential-energy barriers to rotation and translation. By lowering activation energies, the local defect induced by the fifth neighbour is therefore acting as a catalyst.

Phillip Ball



A snapshot from Sciortino *et al.*'s simulations, showing the network of hydrogen bonds and, superimposed, a tetrahedral cluster (left) and a five-coordinate cluster with two molecules (top) coordinated to a single hydrogen atom.

The eastern arm of the East Africa rift system, which includes the Kenya rift, is regarded as the exemplary area for active asthenospheric involvement in rifting. But our near-surface data, revealing strong asymmetry in the distribution of rift-basin faults and off-rift volcanism, apparently indicate that the lithospheric extension is occurring by simple shear — with a component of rotation and a large lateral offset of deformation with depth — so ruling out a role for the asthenosphere immediately beneath the surface rift. The new data now suggest that this asymmetry is merely skin deep. The occurrence of the seismic anomalies in the lower crust and upper mantle vertically below the rift represents convincing evidence of a pure-shear mode, without rotation, of extension in the lithosphere beneath the Kenya rift. And the apparently diapiric intrusive shapes of the upper mantle seismic velocity anomalies strongly suggests an active role for the asthenosphere in the rifting.

The location of the Kenya rift near a major crustal boundary could indicate that weakness or stress-concentration in the lithosphere at this transition localized the rifting. But the boundary could also mark a deep transition in lithospheric thickness which may localize active intrusion of asthenosphere into the lithosphere. Although the new data provide tantalizing clues, the ultimate roles of active or passive asthenospheric processes in rifting are probably interrelated⁹.

The new data indicate that the mantle portion of the lithosphere is almost completely separated beneath sections of the Kenya rift, a condition necessary for successful rifting and the formation of a new ocean. If the regional stress conditions are right, we could expect to see, ultimately, the birth of a new East African Ocean. As Africa is surrounded on the east and west by spreading ocean ridges, however, and rifting in the Red Sea and Gulf of Suez has succeeded, the Kenya rift may be doomed to failure. The active component of rifting looks successful, but the passive stress field does not. □

Paul Morgan is in the Department of Geology, Box 6030, Northern Arizona University, Flagstaff, Arizona 86011-6030, USA.

Nucleons just yield to pressure

Barry R. Holstein

ONE of the important lessons that particle physicists have learned over the past two decades is that most properties of fundamental particles can be understood in terms of the simple quark model. In this picture, the proton (neutron) is viewed as a bound state of three point-like quarks — two (one) 'up' quarks having 2/3 of the proton charge and one (two) 'down' quarks having -1/3 of the proton charge. But two new papers in *Physical Review Letters*^{1,2}, which measure the proton's and neutron's polarizability, challenge this elementary picture.

The electric and magnetic polarizability of a system measures its deformation when placed in an external electric or magnetic field. Under the influence of an electric field the positive and negative charges of an atom will move in opposite directions. The product of the charge and the resulting separation is the electric dipole moment, and the electric polarizability α_E represents the constant of proportionality between this induced moment and the applied electric field. Polarizability then provides a measure of the strength of the binding between the charges — weakly bound systems have large polarizabilities.

Similarly, in a magnetic field, positive and negative charges move in opposite circular trajectories leading to a magnetic dipole moment. The constant of proportionality between the induced moment and the applied field is the magnetic polarizability β_M . The units of polarizability (electric and magnetic) are those of volume and a very polarizable system has values of the order of the geometric volume — in the case of the nucleon (a proton or neutron) 10^{-39} cm^3 . But experimentally, the size determined for the nucleon polarizability is only a thousandth of this geometric volume, implying that the quarks are tightly bound. Consequently, large fields are required to produce a measurable effect, and in the case of the proton this need was addressed by a University of Illinois collaboration by using 50-MeV γ -ray photons as the source of field, during the process of photon-proton (Compton) scattering¹. Unfortunately there is an important background signal due to classical (Thomson) scattering, without polarization, from the proton charge, which occurs at a level an order of magnitude higher than that associated with polarizability. So high precision is required to separate the uninteresting Thomson from the interesting structure-dependent scattering. The collaboration achieved this by using so-called tagged photons: the photons are radiated from

deflected electrons and each is tagged by detecting its associated electron. This technique has a cost in terms of low event rate, but the Illinois group obtained a measurement accurate to 2 per cent of the photon-proton cross-section, enabling the first relatively precise determination of the proton polarizabilities: $\alpha_E = (7.0 \pm 2.5) \times 10^{-43} \text{ cm}^3$; $\beta_M = (3.3 \pm 2.5) \times 10^{-43} \text{ cm}^3$.

In the case of the neutron, as there is no net charge, scattering associated with the polarizability is not obscured by the Thomson process. But the Compton scattering measurement is impossible because there is no such thing as a neutron target. A second collaboration solved this problem² by exploiting a second-order process (resembling the atomic van der Waals attraction) whereby the presence of an electric field polarizes the neutron, which then acts back on the field source. This induced effect leads to an interaction between a neutron beam and a charged target which is long-ranged, decreasing as the fourth power of the distance. The collaboration used a pulsed neutron source at Oak Ridge National Laboratory, scattering neutrons from a ^{208}Pb target. Isolating the small polarizability effect from the dominant neutron-lead nuclear interaction is achieved by seeking a unique, characteristic linear term in the energy dependence of the scattering. A precise measurement of the energy dependence of the total reaction cross-section results in values for neutron polarizability comparable in precision to those found for the proton: $\alpha_E = (12.0 \pm 2.5) \times 10^{-43} \text{ cm}^3$; $\beta_M = (3.1 \pm 2.5) \times 10^{-43} \text{ cm}^3$.

These measurements reveal the electric polarizability of the neutron to be possibly 50 per cent larger than that of the proton. If confirmed, this result would be something of a surprise to theorists, as it contradicts the simple (symmetric) quark model prediction requiring the neutron and proton electric polarizabilities to be equal. Of course, a naive quark picture is not completely accurate as it also predicts for the neutron an exact cancellation between the positive (up quark) and negative (down quark) charge distributions within the neutron, whereas experiment reveals a net residue of negative charge near the surface. But even when this is accounted for by inclusion of an interaction between quark spins, the predicted equality of neutron and proton polarizability survives. A simple (symmetric) three-quark model is unable to explain such a sizable difference in polarizability. In this regard, the conclusion is similar to

1. Verney Green, W., Achauer, U. & Meyer, R. P. *Nature* **354**, 199–203 (1991).
2. KRISP Working Group *Nature* **354**, 223–227 (1991).
3. Buck, W. R. et al. *Tectonics* **7**, 213–234 (1988).
4. Neumann, E.-R. & Ramberg, I. B. *Tectonics and Geophysics of Continental Rifts* (eds Ramberg, I. B. & Neumann, E.-R.) 409–424 (Reidel, Dordrecht, 1978).
5. Sengör, A. M. C. & Burke, K. *Geophys. Res. Lett.* **5**, 419–421 (1978).
6. KRISP Working Group *Nature* **325**, 239–242 (1987).
7. Dahlheim, H.-A., Davis, P. & Achauer, U. *J. African Earth Sci.* **8**, 461–470 (1989).
8. Savage, J. E. G. & Long, R. E. *Geophys. J. R. astr. Soc.* **82**, 461–477 (1985).
9. Morgan, P. & Baker, B. H. *Tectonophysics* **94**, 1–10 (1983).

that from recent measurements of proton spin content, also inconsistent with the elementary picture³.

Despite this dilemma and the shortcoming of the simple quark picture, neither the polarizability nor the spin-content measurements can be said to reveal violation of any fundamental principle, and either can be understood by invoking more sophisticated models. Another indication that there is no violation of fundamental tenets is found in a theoretical paper which appeared as a companion to the Illinois measurement⁴. The authors undertake a calculation of the polarizability based on a property of quark interactions called chiral symmetry. The modern picture of the quark-quark force is in terms of quantum chromodynamics (QCD) wherein quarks interact via exchange of massless but strongly interacting gluons. If quarks were themselves massless, QCD would have an exact symmetry under interchange of up and down quarks that have their spin oriented parallel (right-handed) to their momentum and under separate interchange of up and down quarks with their spin antiparallel (left-handed) to their momentum.

This invariance, chiral (handedness) symmetry, is broken (slightly) in nature by the small but finite quark masses. But although the consequences of chiral symmetry are straightforwardly represented in terms of quark-gluon degrees of freedom, the corresponding implications for interactions of bound states such as protons and neutrons and other physical particles are far more subtle. It has become realized how such chiral implications can be represented in terms of an effective expansion in powers of momentum which is valid up to moderate energies and these techniques have been used in elucidating several low-energy phenomena. In particular, Bernard *et al.*⁴ used chiral techniques to evaluate the proton and neutron polarizabilities, a herculean task, involving the evaluation of 52 separate integrals in the case of the proton. Moreover, many of the separate components of the calculation are formally infinite and must be defined by use of some sort of regularization procedure.

Nevertheless the sum of all the integrals is finite and yields values for the neutron and proton electric polarizabilities in reasonable agreement with those found experimentally. For magnetic polarizability, the calculation is not as successful, yielding values opposite in sign to those observed experimentally,

but this is not unexpected as the calculation omitted the (positive) contribution from an important high-energy state.

Clearly, when elementary particles are examined with the precision now possible, the simple models of the past are no longer able to account for some of the new features found. This richness indicates once again that nature is far more

subtle than simple pictures of particle structure have allowed for — subtle but, as shown by the chiral symmetry calculations, not malicious. □

Barry R. Holstein is in the Department of Physics and Astronomy, University of Massachusetts, Amherst, Massachusetts 01003, USA.

EVOLUTIONARY AND DEVELOPMENTAL BIOLOGY

More fecund but not so fit

H. C. J. Godfray and Paul H. Harvey

POPULATIONS of the soil nematode *Caenorhabditis elegans* normally consist almost exclusively of self-fertilizing hermaphrodites. The animal first matures about 300 sperm and then a much larger number of oocytes (eggs). Nearly every sperm is used to fertilize an egg and so the maximum fecundity is around 300. Why doesn't the nematode mature more sperm and thus increase its fecundity? In a paper in the *Proceedings of the Royal Society* (B246, 19–24; 1991), J. Hodgkin



C. elegans, a worm turned to trade-offs.

and T. M. Barnes provide both an elegant answer and a rare insight into the mechanistic basis of an important life-history trade-off.

Mutations that lead to an increase in fecundity may not be favoured by natural selection if they have adverse effects on other aspects of an organism's fitness. In the jargon of life-history theory, there may be trade-offs between increased fecundity and such traits as survival or the rate of development. Trade-offs are at the heart of life-history theory, yet are hard to detect and characterize, a frequent problem being the lack of understanding of the physiological and genetic constraints that determine their form.

A variety of mutations influence allocation by the nematode to male and female function; some have gross effects, causing complete masculinization or feminization, whereas the effects of others are more subtle. In one mutant, *tra-3(e2333)*, the number of sperm produced by the hermaphrodite is increased, so increasing fecundity. Self-

fertilizing individuals of a standard strain produce a mean of 327 fertilized eggs (standard deviation 28), whereas individuals that are genetically identical except for being homozygous for *tra-3(e2333)* produce a mean of 499 (standard deviation 69). The *tra-3* gene has a regulatory function and homozygous mutants seem to be anatomically and behaviourally indistinguishable from the reference strain, apart from an occasional defect in egg-laying which is only expressed in a minority of individuals. At first sight, then, *tra-3(e2333)* would seem to be the type of rare, beneficial mutation that would be favoured by natural selection.

To compare the overall fitness of wild-type and mutant nematodes, Hodgkin and Barnes estimated the rates of increase of populations of the two genotypes. Their assay involved measuring the length of time required for a population of nematodes to exhaust a bacterial lawn grown on agar. Despite the mutant having a higher fecundity, the population of wild-type nematodes grew more quickly and thus had a higher fitness.

It turns out that the mutant fails to outpace the wild type because there is a trade-off between increased fecundity and average generation time. In wild-type individuals, 19.3 (± 0.5) hours elapses between the time of entry into the last larval instar, and the time that the first larval progeny hatches; in mutants the equivalent time is 21.9 (± 1.1) hours. The mutant requires about 2.5 hours to mature the extra 200 or so sperm, and this delays the onset of oogenesis. In a geometrically growing population, progeny produced early in life are at a great premium compared with progeny produced later: the sooner you reproduce, the sooner your offspring can themselves reproduce. The lost 2.5 hours is enough to outweigh an increase in fecundity of more than 50 per cent.

A feature of the biology of *C. elegans* is that the male gametes mature first, followed by the female. The critical life-history decision is thus when the switch occurs. A second, and currently unanswered question, is why gameto-

1. Federspiel, F. J. *et al.* *Phys. Rev. Lett.* **67**, 1511–1514 (1991).

2. Schmiedmayer, J. *et al.* *Phys. Rev. Lett.* **66**, 1015–1018 (1991).

3. Roberts, R. G. *Nature* **344**, 708–709 (1990).

4. Bernard, V. *et al.* *Phys. Rev. Lett.* **67**, 1515–1518 (1991).

RÉSUMÉ

genesis is organized in this way. A likely answer is that sequential production is more efficient than concurrent production (it would be interesting to discover a developmental mutant that allowed the simultaneous production of eggs and sperm). It can also be asked whether a further reduction in sperm production might be favoured by natural selection if this led to a greater decrease in generation time. A temperature-sensitive mutant of the *fem-3* gene (*e2006*) markedly reduces sperm production and leads to a modest decrease in generation time. However, this mutant has a much lower fitness than either the wild-type or the *tra-3(e2333)* genotype. So far, no mutant has been identified that outperforms the wild type.

The use of specific developmental mutants as a tool to explore trait fitnesses is a notable development in life-history studies. The work also provides an excellent example of a trade-off — between fecundity and development time — that is often invoked though little demonstrated. Classical life-history theory normally predicts that selection for high fecundity will be greatest in

r-selected populations continually kept below carrying capacity by density-independent mortality. One curious consequence of the natural history of *C. elegans* is that in crowded populations one would expect natural selection to favour genes for higher fecundity; the advantage of a short generation time shows itself only when the population is exponentially growing most of the time.

Hodgkin and Barnes's study raises other questions. Why, for instance, does the nematode produce many more eggs than it has sperm to fertilize? Presumably these extra eggs are occasionally fertilized by rare males. But in that case, why are there not more males in the population? Studies with *C. elegans* have proved highly productive in several fields; they now seem set to build bridges between evolutionary and developmental biology. □

H. C. J. Godfray is in the Department of Biology and the NERC Centre for Population Biology, Imperial College, Silwood Park, Ascot SL5 7PY, UK. Paul H. Harvey is in the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

ELECTRON TRANSFER

Switches in enzymes

Barry Halliwell

LIVING organisms move electrons over considerable distances, with great speed and specificity: indeed, it is the flow of electrons through the metabolic network that sustains all life. An appropriate arrangement of transition-metal centres can mediate electron transfer because these metals readily accept and donate electrons, but in biological systems such reduction-oxidation centres are often insulated from one another by lumps of protein. How, then, do electrons cross this barrier? J. E. Baldwin, G. M. Morris and W. G. Richards¹ describe a neat switching mechanism that they say can do the job.

The complex metalloprotein systems of respiration and photosynthesis are essential to energy production. The superfamily of cytochromes P-450 are haem-containing enzymes that also use electron flow to accomplish a wide range of tasks, from hydroxylating toxic compounds to encourage their secretion, to participating in the synthesis of hormones and bile acids¹. Finding out how electrons are transported is also important for understanding why these processes sometimes go wrong: if electrons

leak away from their correct path between reduction-oxidation centres, or if too many are passed on at a time, harmful free oxygen radicals can result that may be involved in the ageing process and several human diseases, including cancer^{2,3}. So how do electrons get from the surface of a protein to an internal metal ion-containing centre? In enzymes containing two or more spatially separated centres, how are electrons transferred between them? In multi-component electron-transfer systems, what makes the electrons follow the correct path rather than short-circuiting^{4,5}?

Baldwin *et al.* use a cytochrome P-450 system from the soil bacterium *Pseudo-*

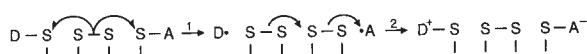


FIG. 1 Electron transfer by covalent switching requires spin unpairing and the breaking of covalent bonds in a heterolytic fashion (both electrons going to one of the products). The example shown here involves sulphur ligands and disulphide bridges. D is an electron donor and A an electron acceptor. In step (1) single-electron transfers occur from the disulphide bridges to the other two sulphurs, generating D• and A• radicals. In step (2) a double-electron transfer regenerates the disulphide bridge. The net effect is electron transfer from D to A. This mechanism is postulated to mediate electron transfer from the reduced FAD ring on putidaredoxin reductase to the iron-sulphur cluster of oxidized putidaredoxin via an intermediate disulphide. (Figure reproduced from ref. 1.)

Small change

ONE of the classic physics experiments of this century has been reincarnated by radioastronomers who measured the deflection of radio waves from the distant galaxy P0201+113 as they passed close to Jupiter (*Astr. J.* **102**, 1879–1888; 1991). The classic experiment was by Arthur Eddington and colleagues, showed who during a total eclipse in 1919 that the curvature of space by the Sun's gravity, predicted by Einstein, was sufficient to deflect the light from a distant star. But where Eddington and colleagues sought a deflection of 1.75 arcseconds, Jupiter's deflection of just 17 milliarcseconds required R. N. Treuhaft and S. T. Lowe to combine radiotelescopes in California and Australia (separation 10,600 km) to give astrometric measurements with the remarkable resolution of 160 microarcseconds.

Yeast released

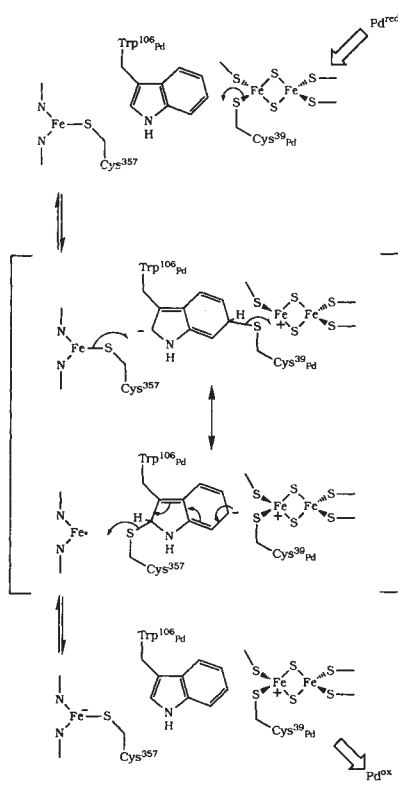
WHAT does yeast want with an enzyme that digests chitin, the cellulosic bathmat that makes up certain cell walls? This chitinase has now been cloned and sequenced by M. J. Kuranda and P. W. Robbins (*J. biol. chem.* **266**, 19,758–19,767; 1991) and has a curious structure; first comes a signal sequence; then a region that contains the catalytic site; and, separated from this by an abundantly glycosylated segment, in which more than half the residues are serine or threonine, is a chitin-binding domain to fasten the enzyme to its substrate. And the need for the chitinase? Kuranda and Robbins have the answer, for they show that when the gene is wrecked, the dividing yeast cells form indissoluble clusters. Now yeast cell walls contain only about 1 per cent of chitin, but all of it at the septum that divides mother from daughter cell — a kind of umbilical that only the chitinase will cut.

Current change

JUST as canaries were used to warn of gas in mines, brachiopods could be used as indicators of climate change in the geological record say G. B. Curry and K. Endo (*Geology* **19**, 1101–1103; 1991). Over the past 10,000 years, *Terebratulina retusa* has spread from the Canary Islands as far as Spitzbergen; the cold-loving *T. septentrionalis*, by contrast, thrives in the Davis Strait between Greenland and Baffin Island, venturing as far south as eastern Canada, their larvae borne by the Labrador current. But isolated relict populations still survive off Finnmark and Norway. It seems that the end of the Ice Age signalled a rapid northward migration of *T. retusa* at the expense of its cousin, and that the dispersal of brachiopod larvae could be used to chart ocean currents in the more distant past.

Metabolic tuning

FIG. 2 Electron transfer from putidaredoxin to cytochrome P-450 by covalent switching (adapted from ref. 1). This second type of electron transfer involves intermediate aromatic amino-acid residues (Trp, Phe or Tyr), which can allow electron transfer between sulphur atoms too far apart to interact directly. The binding of reduced putidaredoxin (Pd) to ferric P-450 is suggested¹ to be accompanied by a conformational change within the putidaredoxin that moves its C-terminal tryptophan residue close enough to Cys 39 (attached to the Fe_2S_2 cluster) to permit interaction. An ion-pair is produced that has the correct configuration to allow covalent switching with the haem-Fe-Cys³⁵⁷ sulphur of P-450, breaking the haem-Cys³⁵⁷ bond and resulting in a briefly tetracoordinate haem iron. The second stage of covalent switching transfers Cys³⁵⁷ from tryptophan back to haem to reform the Cys³⁵⁷-haem bond, and results in overall electron transfer from putidaredoxin to the haem ring. Oxygen then binds to the haem. A second electron is delivered to the haem from reduced putidaredoxin, presumably by the same covalent switching mechanism involving tryptophan. Cleavage of the peroxo-bond may occur simultaneously, generating the substrate-hydroxylating haem-oxo-iron complex. The covalent switching breaks the Cys³⁵⁷-haem bond (the Cys³⁵⁷ is temporarily attached to Trp¹⁰⁶ of putidaredoxin). This may be advantageous in preventing the oxo-haem from oxidizing this sulphur ligand — in the terms of Baldwin *et al.*¹, the switch is insulational (switch on: electrons flow to haem; switch off: haem is insulated). Not until the oxo-iron has hydroxylated the substrate is the haem-Cys³⁵⁷ bond restored.



monas putida to illustrate the principles that may be involved. It turns out that the protein components of the enzyme are not just obstacles, but that they help to control the specificity and rate of electron transfer⁴ by mechanisms involving sulphur ligands, disulphide bridges and aromatic amino-acid residues^{1,4}. In *P. putida* the cytochrome P-450 receives electrons from putidaredoxin (a ferredoxin with an Fe_2S_2 cluster; see Fig. 1), which in turn is reduced at the expense of NADH by a flavoprotein enzyme, the FAD-containing putidaredoxin reductase. The reduced P-450 uses bound oxygen to hydroxylate camphor, a first step in the use of this substrate for energy production. Baldwin *et al.* claim that the mechanism they propose can account for the entire electron transport from NADH through to the cytochrome. The intermediate aromatic amino-acid residues crucial to the scheme of covalent switching (Fig. 2) are conserved among the many cytochromes whose sequence is known; most eukaryotic cytochromes P-450 have specially evolved their own intramolecular tryptophan electron-transfer mediator. The covalent switch involves the redistribution of electron pairs through a framework of momentarily fixed nuclei, so is linked to the nuclear geometry of the molecular system. Baldwin *et al.*

were able to use the high-resolution crystal structure of cytochrome P-450 bound to its camphor substrate, but had to model the putidaredoxin by sequence homology with a ferredoxin, whose crystal structure is known.

The covalent switching mechanism can convincingly explain the details of the organic reaction, and is consistent with the protein structural considerations and the genetic diversity of the cytochromes P-450, in which a high specific activity has been sacrificed for a broader range of catalytic abilities. Baldwin and colleagues suggest that their scheme could have implications for organic room-temperature superconductors, and for the use of organometallic systems designed to extract electrons from organic conductors. Certainly, although some of the chemistry described is not particularly new⁴, the proposed mechanism is both plausible and provocative. □

Barry Halliwell is in the Division of Pulmonary-Critical Care Medicine, University of California Davis Medical Center, Sacramento, California 95817, USA.

1. Baldwin, J. E., Morris, G. M. & Richards, W. G. *Proc. R. Soc. Lond. B* **245**, 43–51 (1991).
2. Halliwell, B. & Gutteridge, J. M. C. *Free Radicals in Biology and Medicine* (Clarendon, Oxford, 1989).
3. Ames, B. N. *Free Rad. Res. Commun.* **7**, 121–128 (1989).
4. Farver, O. & Pecht, I. *FASEB J.* **5**, 2554–2559 (1991).
5. Thomson, A. J. *Nature* **350**, 22–23 (1991).

LAST week Daedalus was contemplating reaction cycles, such as the Krebs cycle which oxidizes glucose in the body. He decided that the repetitive molecular expansion and contraction produced by such a cycle must generate a weak local sound at each site of the reaction. A strong external note at just that frequency could pull all the reaction sites in a sample of tissue into synchrony. Thereafter, the tissue would continue to sing coherently at its Krebs frequency, staying in synchrony by self-reinforcement.

Daedalus is now using this principle to control the rates of cyclic reactions. An external sonic frequency entrains a biochemical cycle by locking the molecular expansion part of the reaction into the low-pressure phase of the frequency, while the molecular contraction part is concentrated in the high-pressure phase. If the frequency is then gradually increased, the cycle will speed up to follow it; it might even reach double speed before falling out of lock with the driving frequency. Similarly, lowering the driving frequency must slow the reaction.

At a stroke, biochemists and physiologists will acquire wonderful new powers. A simple sonic driver, tuned to the crucial cyclic biochemical pathways, could double the output of industrial fermenters and brewery vats. A second driver could slow down unwanted side-reactions or sabotage interfering organisms. Crops and vegetables could be given a cunning acoustic boost — those gardeners who sing to their plants may already be operating in just this way. Medicine, too, will be transformed. Victims of exhaustion will be sung back to life by speeding their Krebs oxidation of injected glucose; parasites and infectious bacteria will be killed or weakened by slowing their metabolic rate.

Even better, this powerful treatment can be applied locally. Slimmers will welcome the DREADCO acoustic vibromassager, whose piezoelectric contact pad will magically melt their unwanted fatty tissue by speeding up its fatty-acid oxidation cycle. And the DREADCO biocyclic hat will give wonderful selective control of the brain. With its aid, poets will speed the metabolism of their creative right hemisphere, while artists selectively galvanize their visual cortex and acrobats 'tone up' their sense of balance and spatial awareness. Manics and depressives will gladly control their aberrant dopamine and noradrenaline brain chemistry. And we will all welcome the ability to slow the whole cortex into perfect, dreamless sleep: from which a smooth rise in frequency can awaken us refreshed in the morning.

David Jones

Efficiency of climate policy

SIR — There is growing interest among policymakers about the cost of alternative approaches for reducing emissions of carbon dioxide. Particular attention has been paid to the use of energy taxes to limit emissions.

Our analyses^{1,2} suggest that the problem of designing an efficient tax system to reduce CO₂ emissions is more complicated than indicated in the literature. Although existing studies have attempted to bound the likely range of taxes necessary to limit CO₂ emissions and the resulting macroeconomic impacts, little consideration has been made of variations in the cost and effectiveness of taxes as they are imposed at different levels of the market. For example, taxes may be imposed at primary production (for example, the well-head or mine mouth), or at the distribution to end-use level (for example, natural gas delivered to homes for space heating; oil deliveries to utilities to produce electricity).

To illustrate the importance of this issue, we used the Environmental Protection Agency's GEMINI model to examine the effectiveness of taxes on carbon and the energy content of fuels, and *ad valorem* taxes, to stabilize CO₂ emissions in the United States at 1990 levels, when they are imposed at different levels of the market. The results indicate that smaller carbon and energy-content taxes are needed to stabilize CO₂ emissions when they are imposed at the primary production level than at the end-use level. In addition, the cost to society (measured as the change in consumers' plus producers' surplus) is smaller when carbon and energy content taxes are imposed at primary production. We found that a \$120 per tonne carbon tax imposed at primary production results in stabilization, whereas a higher \$180 per tonne tax is required at the distribution to end-use level. A \$3.3 per million British thermal units (MMBtu) tax imposed at primary production results in stabilization, whereas a \$4 per MMBtu tax is required at the end-use level. The key reason for these results is the same for both types of taxes: the level of the market at which an energy tax is imposed is important when one form of fossil fuel (such as coal) is converted into another fuel (such as synthetic gas). In our analysis, synthetic gas from coal becomes a substitute for domestic and liquid natural gas imports starting in 2010. Synthetic gas used at the end-use level has a much lower carbon content than the coal from which it is converted. Thus, the surcharge due to a carbon tax is lower when the tax is applied to the synthetic gas than when it is applied to the coal at primary production. (Similar

reasoning applies to the Btu tax.)

In contrast, an *ad valorem* tax has the biggest impact on CO₂ emissions when imposed at the distribution to end-use level. A 125% tax imposed at end-use achieves stabilization, whereas no tax less than 200% imposed at primary production stabilizes emissions. Also, the cost to society is smaller when the tax is imposed at the end-use level. This result also arises due to a synthetic fuels effect.

Comparing the results for the three types of taxes examined in this analysis reveals that CO₂ emissions are stabilized at the lowest social cost when a carbon

tax is imposed at the level of primary production. These results, taken together, suggest that unless policymakers consider the level of the market, misleading conclusions could be drawn about the potential cost and effectiveness of climate-change policy.

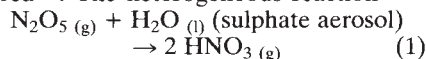
JOEL D. SCHERAGA
NEIL A. LEARY

Office of Policy, Planning and
Evaluation,
Environmental Protection Agency,
Washington, DC 20460, USA

1. Scheraga, J. D., Cohan D., Diener A., Haas, S. & Smith, A. *Gemini: An Energy-Environmental Model of the United States* (EPA, October 1990).
2. Scheraga, J. D. & Leary, N. *Stanford Energy Modeling Forum No 12* (Boulder, Colorado, 27–29 August, 1991).

Future aircraft and global ozone

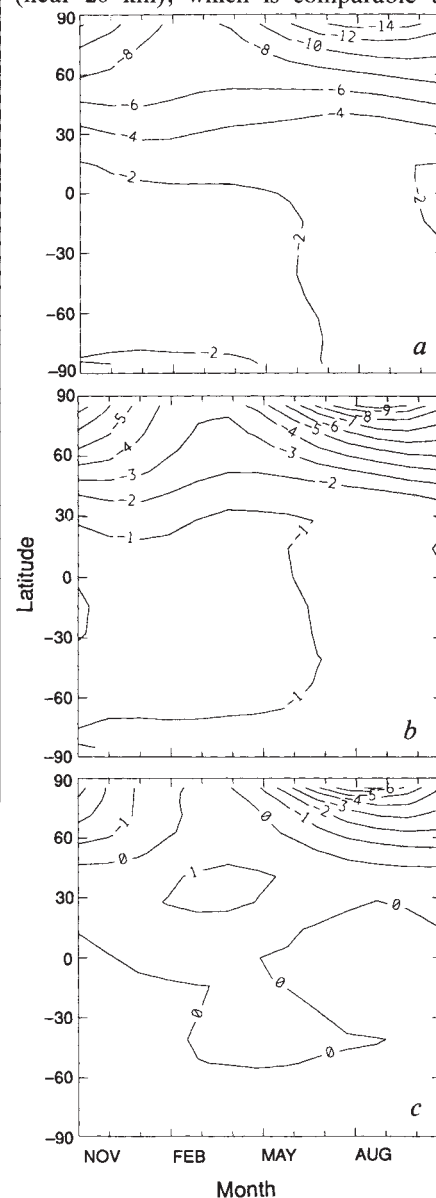
SIR — The impact of the proposed development of commercial fleets of high-speed civil transport aircraft flying in the stratosphere has been studied by Johnston *et al.*¹, who calculated significant global ozone reductions due to NO_x (NO and NO₂) emission, but considered only homogeneous gas-phase chemistry. However, the region of injection coincides with the stratospheric aerosol layer, consisting of droplets of sulphuric acid solution. Heterogeneous reactions on this layer are now thought to affect the global ozone balance², particularly following volcanic eruptions, which considerably increase the available surface area^{3,4}. The heterogeneous reaction



is likely to affect the impact of NO_x emissions by the proposed aircraft, as N₂O₅ is a key reservoir for NO_x in the absence of sunlight.

We have used a radiative-chemical-transport two-dimensional model⁵ to calculate changes of ozone due to NO_x aircraft emission with and without reaction (1). The reaction probability $\gamma = 0.06$ (ref. 6) and the aerosol surface area is varied latitudinally by translating the observed vertical profile of the background surface area³ parallel to the tropopause. The typical hetero-

geneous stratospheric removal time of N₂O₅ is thus calculated to be 6 days (near 20 km), which is comparable to



Calculated column ozone decrease (%) as a function of month for the year 2020 and an injection of 15×10^{10} kg yr⁻¹ (ref. 1) for 5 years (2015–2020) at 19.25 km. NO_x emission index, 15 g (NO₂) per kg (fuel). a, Gas-phase chemistry only; b, including reaction (1) with the background surface area and $\gamma = 0.06$; and c, with doubled background surface area and $\gamma = 0.14$. Calculations are relative to the background atmosphere for the year 2020: Cl_x, 4.8 p.p.b.v.; N₂O, 327 p.p.b.v.; CH₄, 2.4 p.p.m.v.; CO₂, 418 p.p.m.v. We assume a latitudinal distribution of fuel usage: 19° N (9.79%), 28° N (15.64%), 38° N (27.83%), 47° N (33.00%) and at 57° N (13.74%).

loss via photolysis.

Without the inclusion of reaction (1), significant column ozone reductions of up to 14% are predicted (*a* in the figure). The inclusion of (1) has a dramatic effect on calculated ozone depletion. The predicted column changes are approximately halved (*b*).

There is a significant conversion of NO_x to HNO_3 via reaction (1) so that the efficiency of the NO_x catalytic cycle ($\text{NO} + \text{O}_3 \rightarrow \text{NO}_2 + \text{O}_2$ and $\text{NO}_2 + \text{O} \rightarrow \text{NO} + \text{O}_2$) is reduced compared to runs with only gas-phase chemistry. In the model lower stratosphere including reaction (1), the HO_x ($\text{OH} + \text{O}_3 \rightarrow \text{HO}_2 + \text{O}_2$ and $\text{HO}_2 + \text{O}_3 \rightarrow \text{OH} + 2\text{O}_2$) and ClO_x ($\text{Cl} + \text{O}_3 \rightarrow \text{ClO} + \text{O}_2$ and $\text{ClO} + \text{O} \rightarrow \text{Cl} + \text{O}_2$) cycles now dominate ozone loss, so that additions of NO_x have a much smaller effect on ozone abundances.

Because of volcanic eruptions, the aerosol surface area is often well above the background level. We have performed other runs with $\gamma=0.14$ (ref. 7) and doubling the background aerosol surface area. The results show further reductions in the impact of NO_x emissions (*c*). There is even a slight increase (less than 1%) in the column ozone at most latitudes. Injections of NO_2 reduce the efficiency of the HO_x and ClO_x catalytic cycle by the homogeneous formation of HNO_3 and ClONO_2 . The regions of small ozone production are thus due to a decrease in the rate of the HO_x and ClO_x cycle which is not compensated by the increase in the rate of NO_x cycle.

We conclude that the direct impact of high-speed civil transport aircraft on ozone is likely to be much less than previously thought, and that the magnitude of any ozone reductions will critically depend on the sulphate aerosol layer. To improve the modelling of the lower stratosphere, a key area in atmospheric chemistry, more laboratory studies and field measurements are required.

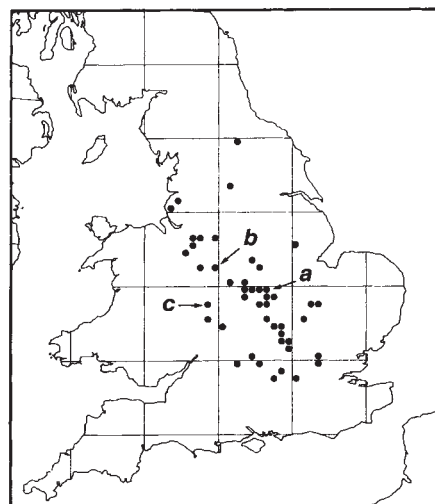
S. BEKKI
R. TOUMI
J. A. PYLE
A. E. JONES

Department of Chemistry,
University of Cambridge,
Lensfield Road,
Cambridge CB2 1EW, UK

1. Johnston, H. S., Kinnison, D. E. & Wuebbles, D. J. *J. geophys. Res.* **94**, 16351–16363 (1989).
2. Rodriguez, J. M., Ko, M. K. W. & Sze, N. D. *Nature* **352**, 134–137 (1991).
3. Brasseur, G. P., Granier, C. & Walters, S. *Nature* **348**, 626–628 (1990).
4. Hofmann, D. J. & Solomon, S. *J. geophys. Res.* **94**, 5029–5061 (1989).
5. Haigh, J. D. & Pyle, J. A. Q. *J. R. met. Soc.* **108**, 551–574 (1982).
6. Van Doren, J. M. et al. *J. chem. Phys.* **95**, 1684–1689 (1991).
7. Mozurkewich, M. & Calvert, J. G. *J. geophys. Res.* **93**, 15889–15896 (1988).

Amphipod also invades Britain

SIR — Following the report in Scientific Correspondence¹ of the invasion of the River Rhine by the immigrant amphipod *Corophium curvispinum* (var *devium*?), I would like to provide an update on the distribution of this species in Britain, where it has also become locally abundant in canals and rivers. First reported



Distribution of *C. curvispinum* var *devium* in the British Isles (with acknowledgements to the Biological Records Centre, Monks Wood) a, Grand Union canal; b, Trent-Mersey canal; c, river Severn.

in 1935 in the River Avon², and later (1970) in the Grand Union canal³, this tubicolous crustacean has been turning up more frequently in nets whenever scrapings of canal revetments are made, particularly in the English midlands (see figure). It is now thought to be an important food for coarse fish and may have been spread in the ballast water or on the hulls of boat traffic⁴.

Van den Brink *et al.*¹ suggested that the amphipod is successful in the Rhine partly because of its toleration of high salinities. Our studies^{5,6} show that *C. curvispinum* is leakier to Na^+ and Cl^- and water than other freshwater amphipod species — for example Na^+ exchange is 18–30% of body content per hour compared to 4–10% in *Gammarus* species⁷. Cladocerans (such as *Daphnia*) have similar sodium turnover rates but, with lower internal Na^+ concentrations, probably expend less energy in replacing ions lost by outward diffusion. A feature of *C. curvispinum* is the physiological variability shown between different populations from different sites, notably in the ability to retain and replace Na^+ and Cl^- .

We believe that the species is becoming better adapted to low salinities (adults can maintain Na^+ balance in as low as 0.5 mM Na^+), and that these differences are indicative of differing degrees of adaptation to new freshwater

habitats. Ion turnover and uptake rates are lowest in a River Severn population (mean annual environmental $\text{Na}^+ = 1.2 \pm 0.1$ mM) and highest in the Grand Union and Trent-Mersey canals. Between 1980 and 1990, sodium permeabilities have decreased at some sites.

Many questions remain as to whether these changes are phenotypic or genotypic, and the picture may be further complicated by multiple introductions into Britain of stock at different times. Nevertheless, the direction of change is clear, towards lower permeability and, therefore, reduced ion turnover, features which *C. curvispinum* shares with other species⁸ and which seem to be important in the evolution of freshwater faunas.

It would be interesting to see if the Rhine populations are as well-adapted to freshwater as British ones. It is probable that the Rhine observations represent the next stage in a westward colonization which started in the Ponto-caspian basin, aided by improvements in river and canal communications. The ancestors of some of the British populations may have been picked up by ships trading with north German ports following the amphipod's appearance in the river Elbe in 1918.

R. R. HARRIS

Department of Zoology,
University of Leicester,
Leicester LE1 7RH, UK

1. Van den Brink, F. W. B. *et al. Nature* **352**, 576 (1991).
2. Crawford, G. I. *Nature* **136**, 385 (1935).
3. Moon, H. P. *Nature* **226**, 976 (1970).
4. Pygott, J. R. & Douglas, S. *Naturalist* **114**, 15–17 (1988).
5. Taylor, P. M. & Harris, R. R. *J. comp. Physiol.* **B156**, 323–328 (1988).
6. Harris, R. R. & Bayliss, D. *J. comp. Physiol.* **B160**, 85–92 (1990).
7. Sutcliffe, D. W. *Comp. Biochem. Physiol.* **52A**, 255–258 (1975).
8. Lockwood, A. P. M., Croghan, P. C. & Sutcliffe, D. W. in *Perspectives in Experimental Biology* (ed. Spencer-Davies, P.) Vol. 1. 93–107 (Pergamon, Oxford, 1976).

Gas sample correction

SIR — P. Allard *et al.* (*Nature* **351**, 387–391; 1991) cited me as having performed an analysis of a gas sample from Paterno on the southern slope of Mount Etna. Not only did I not analyse this sample, the value given by Allard *et al.* is incorrect. Allard *et al.* claim that the CO_2 -rich gas sample had a helium-isotope ratio (R/RA) of 6.6, but my analyses (to be published elsewhere) show that these ratios are 5.9–6.0.

B. MARTY

Laboratoire de Magmatologie et
Géochimie Inorganique et
Experimentale,
Université Pierre et Marie Curie,
4 Place Jussieu,
75252 Paris Cedex 05, France

Captive creatures

Georgina M. Mace

Last Animals at the Zoo. By Colin Tudge. *Hutchinson Radius*: 1991. Pp. 266. £16.99, \$22.

THE past 20 years have seen a dramatic change of emphasis among good zoos. They have shifted from being exhibitors to conservators of wildlife, and from being entertainers to educators. The causes are many and diverse. Changing attitudes among the all-important public to the notion of animals behind bars have certainly had a considerable influence, but much of the impetus has come from within. On the whole, people involved in running zoos are animal-lovers, and the perilous wild status of many species that are well established in captivity has not escaped their notice. There is a genuine desire to help. But for zoos to make a valid contribution to species conservation, more than the odd baby from a captive endangered species is required. A solid foundation in a general conservation movement is needed, as is the development of novel methods in population management, reproductive physiology, husbandry, veterinary and behavioural sciences, as well as wildlife management. In *Last Animals at the Zoo*, Tudge takes us on a whistle-stop tour of the world's zoos, describing projects, ideas and development in conservation breeding. One cannot but be impressed at the wealth of information distilled into this interesting and readable book.

The moral and philosophical questions are important, and Tudge does not duck the challenge. Early in the book, in a 'why bother?' chapter, he discusses in general terms the issue of conservation. Should we really attempt to save all the world's species? If not, should we just save the ones we perceive to be valuable to mankind? Tudge comes down on the side of revering and therefore conserving nature for what it is, accepting that, unless we do, we cannot hope to make the world a better place for ourselves. There is an interesting conflict here though. To attempt to save the world we see today, we have to intervene in the lives of animals; but the intervention is different from that intended to simply benefit our own species. It is not intervention that is the problem for Tudge, it is who that intervention is supposed to benefit. This is a refreshing contrast to the many who believe that almost any-

thing is justified if it is of short-term benefit to mankind.

Tudge then goes on to confront the 'better dead than captive bred' brigade, and counters the idea that captive breeding is simply a way by which zoos can justify their existence to the public. Even under the

K. W. Fink/Ardea
Photographics

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Not left out in the cold — but who is intervention supposed to benefit?

most optimistic schemes for habitat preservation, many wild species will simply not survive the next century. Their numbers in the wild are too small, or their habitats are too fragmented or being too rapidly destroyed by humans. Of course, the species that tend to be most prone to habitat loss are those with the greatest requirements: large-bodied species, species at the top of food chains and those with highly specialized needs. Many of these are actually the kinds of species that are already well known to zoos.

Ultimately, however, captive breeding must lead to re-establishment of populations in the wild. There are a few good examples where this has been achieved, but this was due as much to opportunity as to intent. Already there are populations that cannot be reintroduced because there is no suitable habitat. And timescales are important: in many cases it will be hundreds of years before viable wild populations can be re-established.

Zoos cannot expect to breed all endangered species or to show immediate benefits from breeding the ones that they do, but this does not devalue their essential contribution to conservation. Incidentally, this point is rather at odds with the strangely Victorian-sounding subtitle of the book ("How mass extinction can be stopped"). There is little that zoos can do to stop the mass extinction now under way, although they might do a great deal to reduce its effects.

Zoos have to maintain species without domesticating them so that they retain their potential for adaptive evolution well into the future. This is a large challenge that has two particular repercussions. First, population-genetic theory suggests that captive populations have to be substantially larger than can be incorporated in one or even a handful of zoos. Cooperative programmes involving national or international agreements are essential, and Tudge describes many of the systems of organization that are rapidly developing worldwide to oversee these programmes. Second, reproductive technologies (including the much heralded 'frozen zoo') could be of enormous benefit, as recognized recently by the many groups investigating the reproductive biology of exotic species with a view to managing these animals' reproduction through storage of genetic material. Much of this research is still in its infancy, but Tudge provides an enticing glimpse into a novel and fast-moving field.

Animal welfare is not forgotten in all of this. In particular, those responsible for

zoos are acutely aware of the behavioural effects of captivity, and there are now many ways in which the lives of captive animals are made more interesting. But such 'behavioural enrichment' has wider implications. The animals must be kept in a way that conserves their ability to cope in the wild. Tudge contrasts the differences between reintroducing Arabian oryx to the deserts in Oman and reintroducing golden-lion tamarins to the tropical forests of Brazil. Captively bred tamarins fared poorly — the large component of their behaviour that they learnt in captivity meant they were ill-equipped to cope with the complexities of life in the wild. New approaches to reintroduction were necessary, and have recently been more successful.

Tudge's book is littered with anecdotes and cannot fail to fascinate. I hope that it will show that captive breeding is a serious business. But such breeding will contribute most to conservation if

it is fully integrated into wider plans involving wild populations and their habitats: many of the concerns about maintenance of appropriate genetic traits can be allayed if a population survives in some form in or near its natural habitat. □

Georgina M. Mace is in the Institute of Zoology, Zoological Society of London, Regent's Park, London NW1 4RY, UK.

■ The reintroduction of endangered mammals into the wild is also the subject of *Beyond Captive Breeding*, a volume arising from a symposium held at the Zoological Society of London in 1989. Edited by J. H. W. Gipps, the book contains sections on policy and politics, logistics, and ecology and genetics, ending with several case-studies. Published by Oxford University Press, £45. *Meant to be Wild*, journalist Jan Deblieu reports on biologists' attempts to return North American animals to their natural habitats through captive breeding. Published by Fulcrum, \$24.95. □

Thinly sliced

Philip Rubery

Molecular Activities of Plant Cells: An Introduction to Plant Biochemistry. By John W. Anderson and John Beardall. Blackwell Scientific: 1991. Pp. 384. £49.50, \$99.95 (hbk); £22.50, \$49.95 (pbk).

WHY is an illuminated mesophyll cell like a starving rat? Because it too performs gluconeogenesis, but in miniature. The cytoplasm of the mesophyll cell receives trioses from the chloroplast for conversion to sucrose. In the rat, glycerol from adipose tissue, and lactate and alanine from muscle, are recycled to glucose by the liver. The carbon, whether fixed from the air or freed from reserves, is protected from oxidation and exported as fuel to phloem or blood. Thus might ponder the 'general biochemist' while reading this introductory book, which is principally intended for plant-science students. Its focus is on primary productivity and the construction and renewal of cellular fabric, with emphasis more towards the energetics, mechanism and regulation of processes than the fine detail of structural molecular biology. Although avowedly introductory in scope, the exposition has firm chemical and quantitative foundations and is generally clear, logical and rigorous within its self-imposed limits. The style is not cluttered with latinate polysyllables and circumlocutions that are so often found in current biological writing, and is refreshingly free of jargon. A particular effort is made to use appropriate abbreviations and metabolic termi-

Human life defined, or defied?

S. A. Barnett

Encyclopedia of Human Biology. Volumes 1–8. Editor-in-chief Renato Dulbecco. Academic: 1991. \$1,950, £1,250.

ACCORDING to the art historian Kenneth Clark, authoritarian governments "live by lies and bamboozling abstractions" and therefore dislike encyclopaedias. Clark is referring to Denis Diderot's *Dictionnaire Raisonné* (published in 1751–65), an irreverent work of great historical importance which, in the Age of Enlightenment, was twice censored by the authorities. Today, despite the fact that many of us still enjoy reading and owning books, encyclopaedias are liable to be superseded by discs — floppy or hard. As the editor of this new venture writes, "New information appears daily at . . . an astounding rate". Hence, in the many fields of human biology, even

nology: the occasional lapse, such as calling GTP a trinucleotide, is thus more jolting than usual.

Inevitably the generalities of energetics and molecular machinery — well-trodden ground — are sometimes covered in rather a perfunctory way, but aspects particular to plant structure and metabolism stand out and will prepare students well for further insights — for example, that the complexity of plant cell walls may reflect a signalling as well as a structural role. Perhaps the most useful part of the book for me was the thoughtful treatment of assimilation, with clear and modern accounts of the complexities of C₃ and C₄ mechanisms, of the metabolic traffic between chloroplasts, mitochondria and peroxisomes during photorespiration, and also of nitrogen and sulphur processing. Any hazy preconception of an animal-orientated reader that chloroplasts are simply convenient intracellular milch cows, driven backwards by light, is soon dispelled. Indeed, the regulatory role of light, in metabolism rather than morphogenesis, is a pervasive theme, although I was disappointed that the discussion of control did not place more emphasis on system properties. The authors conclude with a useful account of organelle biogenesis, but their overall treatment of gene expression is too elementary to be free-standing for most types of student.

The metabolic filling in this sandwich has substantial value, and should satisfy the appetites of purchasers who are prepared for their bread to be rather thinly sliced. □

Philip Rubery is in the Department of Biochemistry, University of Cambridge, Cambridge CB2 1QW, UK.

the most authoritative surveys are likely to be soon dated. For this and other reasons, these eight volumes, presented by the publishers as "human life defined", are more likely to be disconcerting to users and reviewers than to tyrants.

The format and illustrations are excellent and encourage browsing; and the whole series has clearly been assembled with great labour and dedication. Articles are longer than those in most encyclopaedias. Each conveniently begins with a glossary and ends with a bibliography.

The glossaries are designed, the editor tells us, as lists of definitions. But the entries in most of the lists are not definitions but notes on words; and some are misleading or confusing, for instance, those for DNA, gene and nucleic acids under "Agronomy" and that for commensalism under "Animal Parasites". In other places too, a firmer editorial hand was needed. The vast topic of human aggression has six references, of which five concern work on monkeys and three are by the author of the article. Elsewhere are other signs of hasty writing. The first words of an interesting contribution with an odd title, "Affective Disorders, Genetic Markers", suggests an unconventional limitation of genetics: "The aim of genetics research is to understand the inherited biochemical basis of the disease state". More important, the crucial phrase "inherited biochemical basis" is not given the critical analysis it requires.

The philosophy behind the choice of subjects is difficult to make out. Of 84 articles in volume 1, 14 are on psychological topics and two ("Altruism" and "Bioethics") are primarily philosophical; these match expectation. But one ("Aqueous Two-Phase Systems") seems to have escaped from a work on physical chemistry. Most of the remainder have a strong medical component. Seven are on bone, from "Bone Cell Genes" to "Bony Pelvis of Archaic *Homo sapiens*"; two are on joints, "Articular Cartilage and the Intervertebral Disc" and "Articulations, Joints between Bones". "Animal Parasites" is followed by a very technical piece (one of many) entitled "Antibiotic Inhibitors of Bacterial Cell Wall Synthesis".

We might justly have expected at least one contribution under Anthropology between the last two. Inspection of the complete table of contents confirms the anthropological shortfall; it also shows a

similar lukewarm interest in human ecology, especially population dynamics; and the treatment of human evolution seems perfunctory.

Some articles in the first volume cover much ground and present important aspects of medicine: a long one on autoimmune disease, one on atherosclerosis, a surprisingly brief one on birth control, and one entitled "Body Fat, Menarche, and Fertility". The contents of others raise the question: who is being addressed? The articles entitled "Adipose Cell" and "Adrenal", both on topics with extensive reverberations in anthropology or psychology, are confined to specialized biochemistry and physiology, and one entitled "AIDS Epidemic (Public Health)" is, astonishingly, entirely about the United States.

The contents of these volumes suggest that, for the editors, human biology is mostly biochemistry, molecular genetics, physiology and clinical medicine, with minor excursions into psychology and other fields. Recent achievements certainly justify expanded interest in molecular and cell biology, but the accompanying greater tendency towards intense specialism is less welcome. Most of the immediate problems of human biology — matters of life and death for our families and our species — concern our interactions with our environment and with ourselves; and the mass of information in these fields requires interdisciplinary treatment.

Correspondingly, the aspects of human biology that arouse the most heated debates are those at the level of whole organisms or of ecology. Contributors to a modern encyclopaedia need not, perhaps, follow the example of the eighteenth-century *philosophes* and take a stand on such matters; but they may reasonably be expected to tell readers of disagreements and difficulties. For instance, an article with another unexpected title, "Behavior: Cooperative, Competitive, and Individualistic", presents approvingly two kinds of interpretation of human action which, many would think, are opposed. One, founded on "Darwinism", reduces humanity to genes or to a product of natural selection. The other is centred on experiments in which subjects are set tasks presenting ethical problems. The article gives no hint of relevant controversy.

No doubt many students, journalists and others, buoyed up by hope of enlightenment, will consult these firmly bound volumes for initial guidance on difficult questions. But not, I fear, for long. □

S. A. Barnett is Emeritus Professor of Zoology in the Australian National University, Box 4, GPO, Canberra, ACT 2601, Australia.

Preparing for the change

Neville J. Woolf

Principles of Adaptive Optics. By Robert K. Tyson. *Academic*: 1991. Pp. 298. \$49.95, £35.50.

TEACHING students about how telescopes form images seems to involve a sequence of correcting for inadequate concepts. We start by teaching geometric optics, suggesting that there is no limit to magnification. We correct this false impression by teaching about wave optics and the diffraction limit. Next, we teach about atmospheric turbulence and the limits it imposes so that even large telescopes produce similar fuzzy images to small telescopes. Then we teach about short exposure and the speckled structure of images that permit the reconstruction of images up to the diffraction limit, but only for a bright-enough source. Now we have a new book that discusses how to correct the wavefront so that sharp images can be obtained up to the diffraction limit even for faint sources.

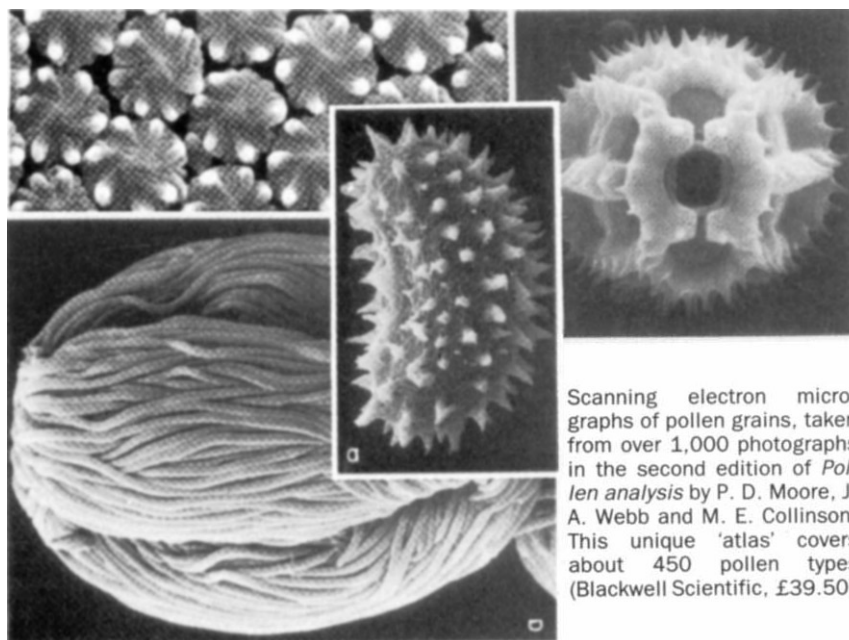
The basic idea is simple. Atmospheric disturbances result in variation in the density of air and thus its refractive index from place to place. And so a distorted wavefront arrives at the telescope. We put a deformable mirror at a pupil plane of the telescope to distort the incident wavefront so that it is reflected back as plane waves. All that is needed is to sense the wavefront disturbance and

appropriately warp the mirror. Unfortunately, the disturbed air is streaming past the telescope. We need in effect to sense the shape of a large mirror and correct it, all in a millisecond, not forgetting that the original shaping took years. And the deformable mirror will not be conjugate with the entire atmosphere, so there will be only a small patch of sky corrected at any one time.

This is the first book on adaptive optics. It is intended to meet the needs of three communities: the astronomers who can afford to build their large telescopes only on the ground, yet who still need diffraction-limited performance; the military who want to know more about objects travelling at high altitude; and other military workers interested in transporting energy by laser beams, presumably to direct them at the same high-altitude objects. It is a strange mixture, ranging from those interested in detecting a few photoelectrons over several hours to those concerned that they do not burn up their equipment in the large amounts of energy involved in their work.

The first impression given by the book is that a student new to the subject has a tremendous amount of material to absorb. But then again, this is an area where the best will eventually extirpate the inferior. Much of what is here will soon be seen as irrelevant or wrong. The worst flaw to my mind is that the optician is not informed that his or her equipment, the telescope, and the building in which it is housed, possibly contribute more to the blurring of the image than does the rest of the atmosphere put together, and that correction of this

Palynological portraits



Scanning electron micrographs of pollen grains, taken from over 1,000 photographs in the second edition of *Pollen analysis* by P. D. Moore, J. A. Webb and M. E. Collinson. This unique 'atlas' covers about 450 pollen types (Blackwell Scientific, £39.50).

source of error is likely to be crucial in determining whether the adaptive system will work. It puts a large dent in any hopes of using existing equipment and facilities with adaptive optics.

The training of most observational astronomers leaves them unprepared for the immense technical skills in design and construction that are required to get to grips with adaptive optics. Early astronomers had to design and build their own equipment because they understood the problems and solutions better than did industry, and could not afford industrial prices. But this certainly gave them an advantage. Similarly, if young astronomers are to take advantage of adaptive optics, their training must change.

Unfortunately, the book will not be very helpful to the trainee astronomer. For example, many have realized that the use of 2-micrometre infrared light is the easy way to become acquainted with adaptive optics, but infrared is not even mentioned in the index. The intended audience of the book is too broad, and too much emphasis is placed on theory rather than practice. And the omission of aspects of speckle will leave the reader without an important tool to understand various options and critical limits of adaptive systems. These are important: for example, when a corrected telescope is between a factor of three and six times too large to be limited by diffraction, the simplest adaptive system changes from being one that does not work to one that is very useful. Nor are the options for guiding on the brightest speckle discussed.

There is a faith in the astronomical community that laser beacons will solve all problems, and that the largest telescopes will soon be operating up to the visible diffraction limit. This is very optimistic. What about the self-spoiling aspects of current telescopes? And should we not consider the time it takes to develop even simple devices, let alone adaptive optics? Finally, laser systems have their limitations in both performance and price. Laser-corrected images will whiz around unless there is a second servoloop to correct this by stabilizing the position of a star image, and these stars will always be faint. Laser beams directed at the atmosphere do not sample the same region of air as the parallel beam from a star, and the problem gets worse as the aperture gets larger. The solution might be to use many different laser beams or starlight, but these add one more layer of complexity. The promise is there, but there will be several more books before the systems are. □

Neville J. Woolf is at the Steward Observatory, University of Arizona, Tucson, Arizona 85721, USA.

Man out of time

Alasdair Whittle

Archaeology: Theories, Methods and Practice. By Colin Renfrew and Paul Bahn. Thames and Hudson: 1991. Pp.543. £18.95, \$29.95.

ARCHAEOLOGY is itself now of some antiquity. It was properly established between 1840 and 1870, when it acquired a timescale and explanations with which to interpret our long past. Since then it has accumulated an astonishing list of achievements. The record of human existence has been pushed back to more than three million years ago. The social and technological success of Palaeolithic people has been charted across the globe. The 'Neolithic revolution', the early postglacial emergence of sedentary lifestyles and food production, has been traced not only in western Asia, but also in eastern Asia, China and the United States. Evidence has also been found around the world of early civilizations, many of them urban, socially complex, literate and technologically advanced, a foretaste of (forewarning to?) the modern world. And the purely archaeological story merges with the ancient and mediaeval worlds known from historical sources. Archaeologists now have lively debates about the nature of historical change, about the evolution of social formations, and about the experience and meaning of being human.

The subject has not lacked for wider appeal beyond its practitioners. Archaeology was already popular in the nineteenth century and has lost none of its intrinsic fascination for many ordinary people. Some sites in England, such as Stonehenge and Avebury, victims of their own appeal, are at risk from too many visitors. And in many Western countries, archaeology has become more or less embedded in the planning process, recognized as an important factor in the modern landscape.

For all this success and popularity, there is still a lot of ignorance about what archaeologists think and do. Many would-be students are attracted by the exotic and glamorous (the Indiana Jones syndrome), and many parents and teachers are still suspicious of the intellectual credentials of the discipline. Even academics have their doubts, from arrogant modern-historians surfeited with information, to scientists mistrustful of what they see as a lack of procedural rigour. And this ignorance can extend to all manner of owners and occupiers of land, as well as to officials, politicians and so on.

Archaeology is doubly satisfying as an introduction to the subject. On the one

hand, it celebrates the achievements of the discipline and revels in the worldwide scope of enquiry and in the vast human timescale. On the other, it takes the reader gently but thoroughly through a veritable encyclopaedia of information about procedures, techniques, explanations and results, enough surely to convince even the most sceptical of the challenge and value of archaeology. That information is well presented by the now-familiar technique of inserts, boxes and special features alongside the regular text, and is intelligently organized around a series of simple questions about people in the past: what did they eat? what did they think? and so on. To its credit, the book does not seek to provide a series of simple answers, although it is always ready to give some idea of what an answer might look like. This should be particularly valuable to university students who are starting the difficult process of thinking about the nature of problems, and who are asked to abandon their desire for certainty.

The book is not only an introduction but also a contribution to the debate about how archaeology should be conducted. In wider terms, it is modernist rather than postmodernist; in archaeology's own unlovely jargon, processual rather than postprocessual. The authors believe in an ultimately knowable past, provided that appropriate procedure, relevant questions and the luck of discovery all coincide. They are not concerned about the relativist interpretation of pasts invented in the present to serve the purposes of the modern world. The authors' position is vigorously argued as a subtext throughout the book, but the opposing view is also given a fair airing. Other modern issues are also discussed, including the disturbance of the dead and the return of cultural materials to the countries of their origin.

It may seem parochial to complain that there is not more on the Mediterranean and Europe, and on the classical and early mediaeval worlds. These are important in many UK university degree courses. The book seems to be aimed at a US market, but the wider geographical scope would be unlikely to do any student much harm.

There have been many distinguished and successful introductions to archaeology. This is, quite simply, by far the best available. □

Alasdair Whittle is in the School of History and Archaeology, University of Wales College of Cardiff, Cardiff CF1 3XU, UK.

Autumn Books

Great Mambo Chicken and the Transhuman Condition, *Junk Science* and *Fantastic Archaeology* are just three of the books to be featured in the Autumn Books supplement in next week's *Nature*. □

A three-dimensional seismic image of the crust and upper mantle beneath the Kenya rift

W. Verney Green*, Ulrich Achauer† & Robert P. Meyer*

* Department of Geology and Geophysics, University of Wisconsin, Madison, Wisconsin 53706 USA

† Geophysics Institute, University of Karlsruhe, D-7500 Karlsruhe, Germany

A tomographic image of the seismic structure beneath the Kenya rift shows its three-dimensional character. In the lower crust, an axial zone of high seismic velocity varies in width and magnitude along the axis, suggesting that the amount of crustal modification also varies along the length of the rift. In the upper mantle, a steep-sided low-velocity body indicates the presence of partial melt, and its shape suggests that diapirs rise from a narrow wedge of asthenosphere beneath the rift axis in response to extension by pure shear.

THE Kenya rift is considered the typical example of a continental rift in the initial stage of continental breakup by extension. Although the rift is divided along strike into independent sub-basins, most studies have assumed the rift is primarily a two-dimensional feature. The relationship between the sub-basins and the lower crustal and mantle structure is not clear. Furthermore, it is not known whether the stresses causing rifting originate in the lithosphere or asthenosphere.

Such issues can be addressed with three-dimensional images created by seismic tomography. Seismic tomography is a technique used to map variations in the propagation velocity of seismic waves, which can then be interpreted in terms of lithology and/or rheology.

Previous seismic studies (Fig. 1a) have found one- and two-dimensional representations of the average structure beneath the rift and in general show thin crust above anomalously low upper-mantle velocities beneath the rift axis¹⁻⁵. The 1985 Kenya Rift International Seismic Project (KRISP 85) teleseismic experiment consisted of two arrays, a 100×110 km two-dimensional array and a 500-km linear array, both of which crossed the central portion of the rift (Fig. 1b). The areal two-dimensional view of travel-time residuals (the difference between observed and theoretical travel times⁶) produced from vertically incident PKP teleseisms⁷ recorded by the two-dimensional array supports the geological observation that sub-basins are markedly asymmetric across strike. The variability in residuals within sub-basins is small along strike, and the residuals decrease between the sub-basins. The two-dimensional cross-sectional image produced using the full range of teleseismic incidence angles (0–30°) recorded by the 500-km linear array shows 8% velocity range in the upper mantle beneath the rift, which is interpreted as 5% partial melt extending up to the Moho⁸. A previous linear teleseismic array (c–c', Fig. 1a) showed an 8% velocity reduction in the upper mantle⁹.

Here we present a three-dimensional compressional-wave velocity model of the crust and upper mantle to 145 km depth, which results from the first application of a two-dimensional array crossing the rift. Our model clearly indicates that the crust and upper-mantle structure are three-dimensional beneath the rift. For example, intrusions of mafic material into the crust are narrow in some places and wide in others. We interpret an axial, steep-sided, low-velocity body in the upper mantle as representing pure shear thinning of the lithosphere, and we find that the

depth to the top of the asthenosphere is variable along the rift axis.

Setting and experiment geometry

The portion of the rift imaged (Fig. 1a) is located at the peak of the Kenya dome, an elliptical region uplifted as much as 1,800 m since the onset of rifting in the Miocene (23 Myr)¹⁰. Bouguer gravity studies show anomalous high-density material in the crust and a low-density zone in the upper mantle¹¹⁻¹⁵. In the central portion of the rift, the valley is 50–70 km wide and bounded by normal faults, with vertical offsets of up to 4 km (ref. 10). The valley floor contains 2–3 km of volcanic and sedimentary fill¹⁰. The rift can be divided into several segments or sub-basins with characteristic lengths of 100–150 km, on the basis of geological observations^{16,17}. It has long been recognized that fault offsets are typically larger on one side of the rift valley, producing asymmetric sub-basins¹⁸. The fault with larger displacement is called the detachment fault. The sub-basins in the study area (Nakuru-Naivasha and Magadi-Natron) have detachment faults on opposite sides, and are reported to be separated by a region of complex deformation termed an 'accommodation zone'¹⁶.

The KRISP 85 teleseismic experiment arrays were positioned to include the area investigated by the KRISP 85 refraction profiles, so that the velocity–depth functions could be used as input for the teleseismic inversion. Furthermore, the 19-site two-dimensional array and five sites from the linear array spanned the axial positive Bouguer gravity anomaly. Average spacing was ~30 km, which means the horizontal resolution is limited to regions 30 km or larger.

Imaging procedure

The detailed three-dimensional image of the compressional-wave velocity structure presented here results from inverting travel-time residuals from 24 teleseismic observation sites and employing the full range of incidence angles and azimuths. Recordings of 79 teleseisms result in 614 travel-time residual observations. We use the robust Aki–Christofferson–Husebye (ACH) method¹⁹⁻²¹ to produce an estimate of the velocity changes from a starting model required to satisfy these residuals. The volume imaged is divided into horizontal layers, which are subdivided into equal-sized blocks with horizontal dimensions roughly equal to the average seismograph spacing. The inversion uses a damped-least-squares solution to seek velocity perturbations (percentage deviations from average layer velocity) for each block that satisfy the travel-time residuals. This solution is used because some regions of the model are well sampled and others are poorly sampled. The resulting model is the best compromise between reducing the variance of the velocity perturbations and minimizing the misfit between the observed data and the predicted arrival times based on the heterogeneous model. There are several possible sources of error in the model: (1) the method causes underestimation of large horizontal velocity heterogeneities by up to 25% (ref. 22); (2) there is some smearing of velocity perturbations between model layers because of the steep ray paths; and (3) errors in the data are mapped into the shallowest and deepest layers. Furthermore, if the block sizes are inappropriate for the geological structure, the resulting

velocity model is a fuzzy picture of the true velocity structure. Nevertheless, the inversion reduces the variance of the travel-time residuals by >75%.

The model consists of two crustal and three mantle layers (Table 1). The bottom of the volume at 145 km depth is chosen on the basis of incidence angles and array dimensions. The initial velocities for the upper three layers are taken from the KRISP 85 refraction results⁵ and an upper-mantle velocity of 8.1 km s^{-1} is used for the two deepest layers. Nine inversions are made, with lateral block boundary shifts of 1/3 block dimension along each horizontal axis from a mean position. Each block is divided into nine smaller equal-sized blocks whose values are the average of the nine inversions, thus reducing the effect of block boundaries.

Tomographic images

Figure 2 shows the results of the inversion. At all depths, a 9–12% compressional-wave velocity change is found crossing the rift (standard errors are in general $\sim 1.0\%$, so velocity changes greater than 2.0% are significant). The resolution matrix, which relates the estimate of the model to the true model, shows better resolution of the central blocks than of the edge blocks. In the lower crust (Fig. 2a), we find a 12% velocity range with the highest velocities beneath the rift axis, suggesting intrusion into the crust of high-velocity material. In the mantle (Fig. 2b–d), there is a 12.5% velocity range with the lowest velocities under the axis, suggesting that partial melt exists at shallow depths. An unexpected result of the inversion is that the variability found along the rift axis is as large as the variability across the axis, which can only be observed by three-dimensional studies.

Cross-sections through the model help show the axial variability. Cross-section A–A' (Fig. 3a), through the Nakuru-Naivasha sub-basin at 0.75°S , shows higher velocity in the lower crust beneath the rift axis and lower velocity beneath the flanks. In the upper mantle, between 35 and 145 km depth, the narrow,

TABLE 1 Input model for ACH inversion

Layer	Velocity (km s^{-1})	Thickness (km)	Block size ($\text{km} \times \text{km}$)	Number of blocks
1. Upper crust	6.0	10	special	23
2. Lower crust	6.5	25	30×30	49
3. Uppermost mantle	7.5	30	30×30	64
4. Second mantle	8.1	40	30×30	81
5. Third mantle	8.1	40	30×30	81

Within the first layer, 'blocks' are limited to the admittance cone encircled by incidence angles $< 30^\circ$.

steep-sided, low-velocity zone seems to dip toward the west. Cross-section B–B' (Fig. 3b), through the accommodation zone at 1.2°S , also shows high velocity in the lower crust beneath the rift, but here the flanks are also underlain by high-velocity material. The steep-sided low-velocity zone is again present in the upper mantle, but it is not as prominent from 35 to 65 km depth as in A–A'. In addition, the upper-mantle structure beneath the accommodation zone is much more axisymmetric than that beneath the sub-basin.

Interpretation of velocity perturbations

How can the 12% velocity range in the lower crust be explained? An appropriate velocity contrast could be produced either by *in situ* phase changes such as dehydration metamorphic reactions driven by elevated temperatures, or by high-velocity mantle-derived material intruded into the crust. Intrusive material could be cooled primary melt such as basaltic dykes or sills. Alternatively, fractional crystallization of parent magma could leave high-velocity material such as gabbro in the lower crust with the remaining liquid fraction rising higher into the

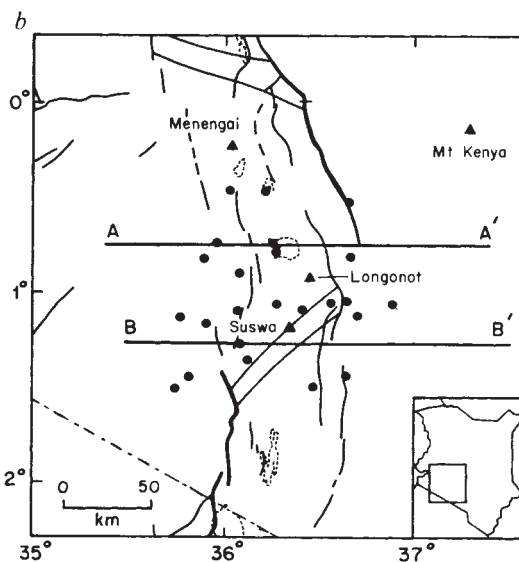
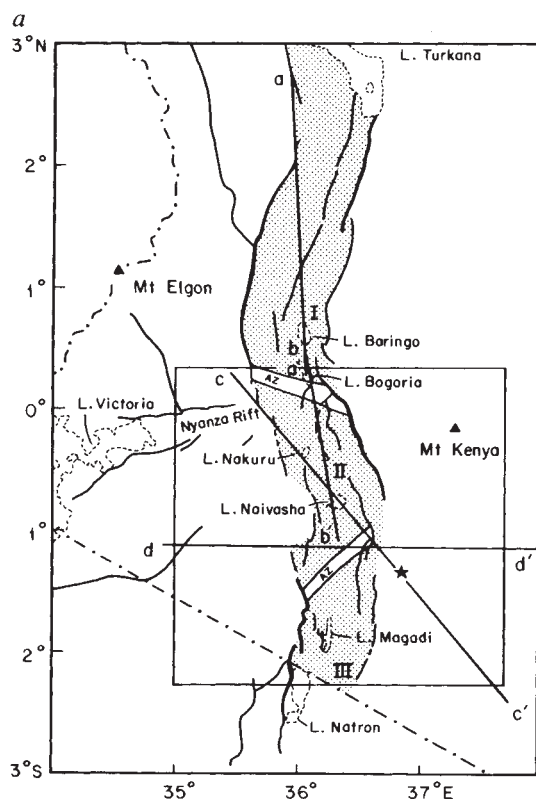


FIG. 1 a, Map of the Kenya rift (stippled pattern). Heavy lines show rift-related faults (after Baker *et al.*¹⁸); the thickest lines are the detachment faults. The thin lines show the locations of previous seismic profiles: a–a' 1968 refraction¹, b–b' 1985 KRISP refraction^{4,5}, c–c' DEAP II teleseismic⁹, d–d' UCLA–Karlsruhe teleseismic⁶. The Kenya rift is divided into three sub-basins: (I) Baringo-Bogoria, (II) Nakuru-Naivasha, (III) Magadi-Natron, separated by accommodation zones (AZ) (after Bosworth *et al.*¹⁵). The box shows the area in b. Star indicates Nairobi. b, The dots indicate teleseismic recording sites. Faults and accommodation zones as in a. Lines A–A' and B–B' show the location of the cross-sections in Fig. 3.

crust or extruding at the surface as alkaline basalts and phonolites. Although the seismic evidence alone is not enough to determine the cause of the high velocity, we favour gabbroic intrusions because isotopic evidence from volcanic rocks supports the fractionation mechanism^{23,24}. It is likely that phase changes have also occurred. The conclusion that intrusion causes the velocity range is supported by results from the KRISP 85 axial refraction profile that found 6.4–6.7 km s⁻¹ lower crustal velocities and a high-velocity layer (7.1 km s⁻¹), which has been interpreted as an intrusion of mafic mantle-derived material^{4,5}.

If the velocity observed beneath the flanks north of 1° S (Fig. 2a) is interpreted as normal crust, the high-velocities observed beneath the rift axis suggest that intrusion is constrained to a narrow zone. South of 1° S, high velocities are observed beneath the rift and flanks, suggesting that the lower crust is much more pervasively intruded. The velocity increases along the axis from north to south, which suggests that either the thickness of the intruded material increases or the percentage of mantle-derived material in the crust increases to the south.

The central volcanoes Mount Susua and Mount Longonot (Fig. 1b) are located above high-velocity regions in the lower crust, as would be expected by fractionation. But the third central volcano, Menengai, is located above low-velocity material in the lower crust. A more detailed study with a tighter seismograph array would decrease the size of resolvable bodies and allow detection of a smaller high-velocity region beneath Menengai, should it exist.

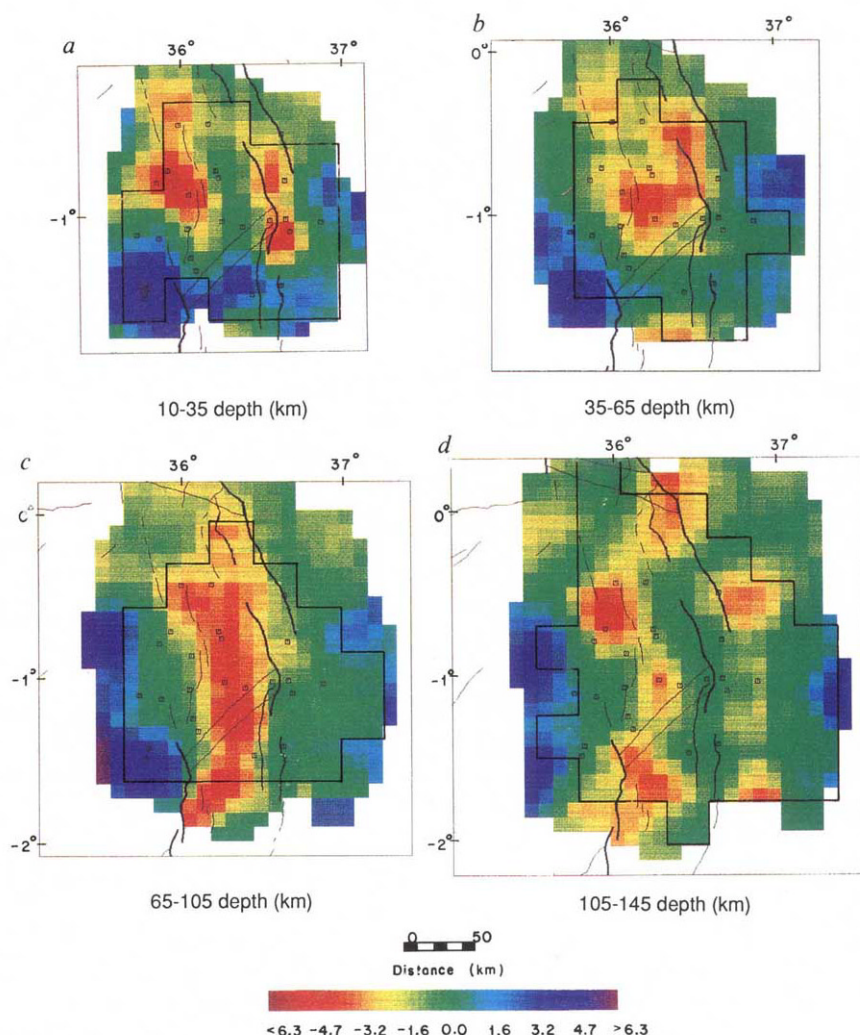
How can the 10–12% velocity range in the upper mantle

beneath the rift be explained? We propose diapirs containing 3–6% partial melt rising from below. Small amounts of partial melt have a large effect on compressional-wave velocity^{25,26} (Fig. 4). If 6% of the velocity range is explained by subsolidus temperatures, the other 6% decrease in velocity could be caused by 3–6% partial melt (Fig. 4b). This partial-melt zone could be either a lozenge (or lozenges) of melted lithosphere or the uppermost asthenosphere. The low velocity persists vertically through the model, suggesting that the partial melt represents asthenosphere. (Although the term asthenosphere is defined in terms of strength, partial melt is likely to have low viscosity and we will therefore refer to zones containing partial melt as asthenosphere).

The lithosphere–asthenosphere boundary is likely to be between 35 and 65 km depth all along the rift axis, which compared with normal shield lithosphere thicknesses of 150–200 km implies lithospheric thinning of 85–165 km. Between 35 and 65 km depth (Fig. 2b), the change in velocity perturbation along the axis of the rift further suggests that the depth to the top of the asthenosphere is variable, being shallowest at ~1° S. The low-velocity zone is steep sided and nearly circular in plan, similar to a diapir.

There is a remarkable correlation between the rift-bounding faults and the low-velocity zone between 65 and 105 km (Fig. 2c), which implies that the lithosphere is thinned directly beneath the rift valley. Furthermore, below 65 km depth, the low-velocity zone becomes more continuous along the axis and broadens with depth, appearing more like a narrow wedge of asthenosphere.

FIG. 2 Inversion results. Colours indicate velocity perturbations (%), that is, differences from average layer velocity rather than actual velocity; therefore, the same colour does not imply the same velocity in different layers, but the same percentage difference in velocity. Red, orange and yellow are slower than average, green is near average, and blue, indigo and violet are faster than average. The box outlines the layer, including unmodelled blocks, in the input model; non-modelled blocks are colourless. The heavy line encloses the area where the diagonal element of the resolution matrix is greater than 0.5. The small squares are station locations. The upper crustal layer (surface to 10 km) is not shown because only the velocity structure of the small-volume of the admittance cone beneath each station is sampled. a, Lower crust, 10 to 35 km; b, uppermost mantle, 35 to 65 km; c, second mantle layer, 65 to 105 km; d, third mantle layer, 105 to 145 km.



The presence of partial melt in the mantle is supported by the regression of PKP residuals against gravity, which showed that the observed change in travel-time residuals associated with the decrease in gravity is five times that of normal sub-solidus conditions⁷. In addition, the Bouguer gravity anomaly derived from a velocity model based on teleseismic profile data and a velocity-density relation is twice as large as the observed Bouguer gravity anomaly²⁷. We note, however, the upper-mantle low-velocity zone observed by the current study is directly beneath the rift axis, whereas that predicted by the PKP average residuals was beneath the west flank⁷. Vertical travel times through the three-dimensional model are consistent with the average PKP residuals, suggesting that the PKP residuals were misinterpreted by not accounting for the depth distribution of anomalous material.

Implications for rifting dynamics

Several hypotheses have been proposed for the underlying causes of lithospheric extension. Models with extensional forces originating in the lithosphere are classified as passive and those with extensional forces originating in the asthenosphere are classified as active²⁸.

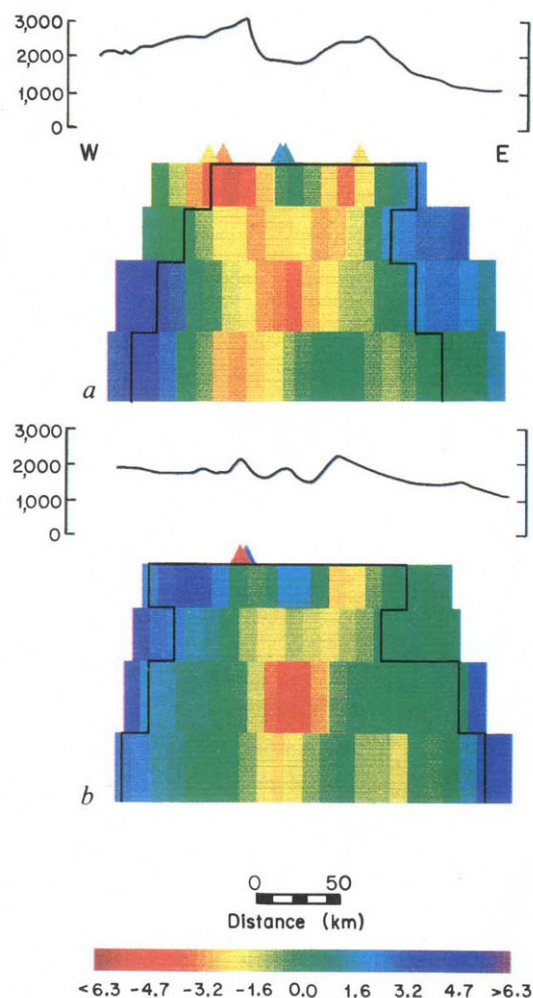
A passive model based on satellite imagery and field mapping suggests that the oppositely facing half grabens in the Kenya rift form by simple shear from tensional forces generated elsewhere in the plate^{16,17}. This model suggests that the detachment fault extends as a fault or shear zone through the entire crust and upper mantle, and that thinning of lithospheric mantle will be offset horizontally 100–150 km from the rift axis. As the

detachment faults occur on alternate sides of the rift axis, the location of lithospheric thinning also alternates. But the thinning observed here is directly beneath the rift axis, in contrast to the predictions of the simple shear model. The results from both two-dimensional profiles^{8,9} show a secondary low-velocity anomaly ~150 km east of the rift, but concur with the finding that the primary location of lithospheric thinning is close to the rift axis.

An active model for the Kenya rift explains rifting as the result of an anomalously hot mantle plume that causes partial melting²⁹. Driven by buoyancy forces, an asthenospheric sheet or 'asthenolith' is forcefully injected into the lithosphere, accompanied by silicic volcanism and mafic underplating of the crust. All extension results from pure shear caused by forceful injection of the asthenolith, which penetrates to within a few kilometres of the surface. This model accounts for the low velocities in the upper mantle, but the observed velocity range in the lower crust and the KRISP 85 refraction velocities are not consistent with asthenosphere just a few kilometres from the surface.

Another model, also a mantle-plume model, explains rifting as the result of a mushroom-head mantle plume, with temperatures 100–200 °C above average, which causes 1.0–2.0 km of uplift³⁰. If thinning of the lithosphere under pure shear follows, caused by the extensional forces of the uplift and aided by gravitational potential, then the asthenosphere passively upwells into the space created by extension. This upwelling causes decompressional melting. The melt is extracted quickly and can be erupted as thick basaltic flows and intruded into the

FIG. 3 Cross-sections through the model and corresponding topographic profiles. As in Fig. 2, colours represent velocity perturbations in % and the heavy line encloses the area where the diagonal element of the resolution matrix is greater than 0.5. The first layer (surface to 10 km) is represented by cones; only those stations within 10 km of the cross section are displayed. The topographic profiles have 20 times vertical exaggeration and the velocity model has no vertical exaggeration. *a*, A–A' at 0.75° S. *b*, B–B' at 1.2° S.



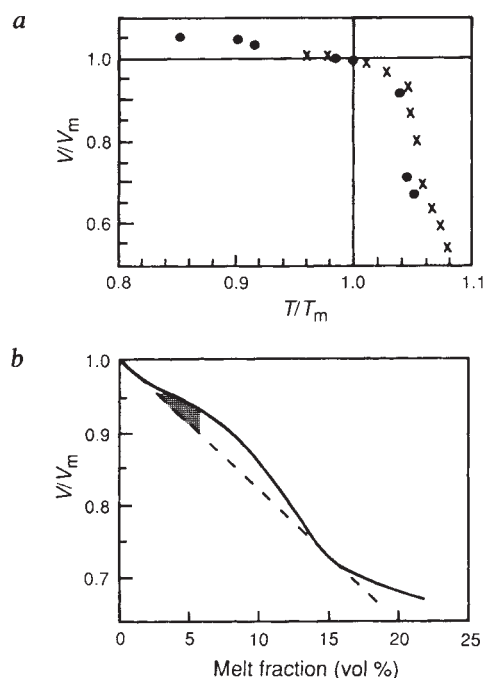


FIG. 4 Laboratory measurements on dry peridotites at pressures up to 1 GPa show the relationship between normalized compressional-wave velocity and homologous temperature and melt fraction. V_m and T_m are solidus velocity and temperature respectively. (After Sato *et al.*²⁶.) *a*, Normalized velocity against temperature. There is a 6% change in velocity at subsolidus temperatures ($T/T_m < 1$), and then a rapid decrease in velocity at higher temperatures. Squares are at 0.5 GPa and circles are at 1.0 GPa. *b*, Normalized velocity against melt fraction. For 10–12% range in velocity, there is a minimum of 3–6% partial melt as shown by the shaded region. (The solid line is from Sato *et al.*²⁶, and the dashed line is from Mavko²⁵.)

thinned crust. During the initial stages, extension may be localized and may not be axisymmetric.

Fluid dynamic and computer modelling place additional constraints on the geodynamics of rifting. Fluid dynamic models show that hot plumes can cause partial melting, uplift and considerable extension, but probably not enough for sustained extension³¹. Computer models show that both active and passive mechanisms could cause lithospheric thinning, but in the active case, lithospheric thinning is confined to a very narrow zone,

and crustal thinning would not occur until the asthenosphere reached the base of the crust³². As the tomographic image shows pure-shear lithospheric thinning confined to a narrow zone, the active case fits well. But the crustal thinning observed by refraction^{4,5} and the asymmetric basins suggest that the passive case cannot be eliminated.

Comparing the results of this teleseismic study with features from the various rifting models, we propose the following rifting mechanism. Rifting results from a combination of passive extensional forces generated within the lithosphere with a hot mantle plume. Decompressional melting causes injection of asthenosphere into the lithosphere driven by buoyancy forces. The asthenospheric injection takes the form of diapirs containing 3–6% partial melt rising from the top of a narrow asthenospheric wedge, which results in the variable depth to the asthenosphere observed along the rift axis. Asymmetric basins develop in the upper crust from external extensional forces, whereas the lower crust and lithospheric mantle thin by pure shear. As 5–8% partial melt is unlikely to remain in the mantle³³, it rises and intrudes into the lower crust, where it undergoes liquid–crystal fractionation. The liquid produces surface volcanism and the more mafic crystalline material remains as high-velocity material in the lower crust. Where the asthenosphere is shallow, intrusion of mantle-derived material into the crust occurs in a narrow zone; where it is deep, intrusion of mantle-derived material into the crust is widespread. This suggests that areas of thinnest lithosphere have well-developed, narrow, magma migration paths, and areas of thicker lithosphere have poorly developed, more diffuse migration paths. The region of the upper mantle containing the highest amount of partial melt is observed directly beneath the rift. At greater depths, however, the low-velocity zone is not always directly beneath the rift, suggesting that the asthenospheric diapirs may not rise strictly vertically or that other deeper diapirs are present. We conclude that although asymmetry is a first-order feature in the shallow structure, it is a second-order feature in the upper mantle. The three-dimensionality of the asthenospheric upwarp and the differences between the velocity structure beneath the sub-basin and the accommodation zone suggests that the segmentation of the rift is related to the spacing and dimensions of asthenospheric diapirs instead of shallow pre-existing structures. Our rifting model is fairly similar to a model based on rift magmatism which has chemically distinct mantle diapirs spaced at regular intervals along the rift axis³⁴. The magmatic model also leads to the conclusion that rift segmentation is caused by the underlying magmatic system rather than shallow pre-existing structures. □

Received 28 May; accepted 30 October 1991.

- Griffiths, D. H., King, R. F., Khan, M. A. & Blundell, D. J. *Nature* **229**, 69–71 (1971).
- Knopoff, L. & Schlue, J. W. *Tectonophysics* **15**, 157–165 (1972).
- Muller, S. & Bonjer, K.-P. *Tectonophysics* **20**, 283–293 (1973).
- KRISP Working Group *Nature* **325**, 239–242 (1987).
- Henry, W. J. *et al.* *Geophys. J. Int.* **100**, 107–130 (1990).
- Herrin, E. *et al.* *Bull. Seism. Soc. Am.* **58**, 1193–1241 (1968).
- Green, W. V. & Meyer, R. P. *Tectonophysics* (in the press).
- Dahlheim, H. A., Davis, P. M. & Achauer, U. *J. Afric. Earth Sci.* **8**, 461–470 (1989).
- Savage, J. E. G. & Long, R. E. *Geophys. J. R. astr. Soc.* **82**, 461–477 (1985).
- Baker, B. H. *Geol. Soc. Lond. Spec. Publ.* **25**, 45–57 (1986).
- Girdler, R. S., Fairhead, J. D., Searle, R. C. & Sowerbutts, W. T. C. *Nature* **224**, 1178–1182 (1969).
- Searle, R. C. *Geophys. J. R. astr. Soc.* **21**, 13–31 (1970).
- Khan, M. A. & Mansfield, J. *Nature* **229**, 72–75 (1971).
- Baker, B. H. & Wohlenberg, J. *Nature* **229**, 538–542 (1971).
- Fairhead, J. D. *Tectonophysics* **30**, 269–298 (1976).
- Bosworth, W., Lambiase, J. & Keisler, R. *EOS Trans. Am. geophys. Un.* **67**, 577–583 (1986).
- Bosworth, W. *Geology* **15**, 397–400 (1987).
- Baker, B. H., Mohr, P. A. & Williams, L. A. *J. Geol. Soc. Am.* **136**, 67 (1972).
- Aki, K., Christofferson, A. & Husebye, E. S. *J. geophys. Res.* **82**, 277–296 (1977).

- Ellsworth, W. L. & Koyanagi, R. Y. *J. geophys. Res.* **82**, 5379–5394 (1977).
- Achauer, U., Greene, L., Evans, J. R. & Iyer, H. M. *J. geophys. Res.* **91**, 13873–13891 (1986).
- Koch, M. *Geophys. J. R. astr. Soc.* **80**, 73–93 (1985).
- Baker, B. H. *Geol. Soc. Spec. Publ.* **30**, 293–311 (1987).
- MacDonald, R. *Geol. Soc. Spec. Publ.* **30**, 313–333 (1987).
- Mavko, G. M. *J. geophys. Res.* **85**, 5173–5189 (1980).
- Sato, H., Sachs, I. S. & Murase, T. *J. geophys. Res.* **94**, 5689–5704 (1989).
- Achauer, U. *Tectonophysics* (in the press).
- Sengör, A. M. C. & Burke, K. *Geophys. Res. Lett.* **5**, 419–421 (1978).
- Mohr, P. *EOS, Trans. Am. geophys. Un.* **68**, 721–736 (1987).
- White, R. & McKenzie, D. P. *J. geophys. Res.* **94**, 7685–7729 (1989).
- Hill, R. I. *Earth & Planet. Sci. Lett.* **104**, 398–416 (1991).
- Keen, C. E. *Geophys. J. R. astr. Soc.* **80**, 95–120 (1985).
- McKenzie, D. P. *Earth planet. Sci. Lett.* **74**, 81–91 (1985).
- Karson, J. A. & Curtis, P. C. *J. Afric. Earth Sci.* **8**, 431–453 (1989).

ACKNOWLEDGEMENTS. We thank the government of Kenya for permission to conduct this research, Leicester University and M. A. Khan for logistical assistance, our colleagues from the University of Wisconsin, UCLA, and the Universities of Karlsruhe and Nairobi for assistance with the field work, analysis or manuscript preparation, and NSF Continental Dynamics programme and DFG Sonderforschungsbereichs for research support. W.V.G. acknowledges the University of Karlsruhe for support.

Minisatellite repeat coding as a digital approach to DNA typing

Alec J. Jeffreys, Annette MacLeod, Keiji Tamaki, David L. Neil & Darren G. Monckton

Department of Genetics, University of Leicester, University Road, Leicester LE1 7RH, UK

Most DNA typing systems used in forensic and legal medicine assay allelic length variation at tandem repetitive DNA regions such as minisatellites. A simple alternative approach that displays patterns of variant repeat units along minisatellite alleles is described here. This produces DNA profiles as extraordinarily variable digital sequences appropriate for forensic investigations, including computer databasing, and for analysing allele diversity and the role of recombination in minisatellite instability.

HYPERVARIABLE human DNA markers have had a profound impact on forensic and legal medicine. Most DNA typing systems use tandem-repetitive minisatellites or VNTR (variable number tandem repeat) loci which can show extreme levels of allelic variability in repeat copy number and therefore DNA fragment length¹⁻⁵. Polymerase chain reaction (PCR) amplification of hypervariable loci, including very short tandem-repetitive 'microsatellites', has greatly increased the sensitivity of DNA typing systems and the ability to type degraded human DNA⁶⁻¹¹. Despite the power of current DNA typing systems, technical problems have prevented their full potential from being realized. Minisatellite allele lengths can vary in a quasicontinuous fashion, making unequivocal allele identification impossible^{2,5,12-17}. Error-prone allele length estimates and electrophoretic 'band-shifts' can occasionally lead to apparent exclusions between 'matching' DNA profiles, and more seriously greatly weaken the potential statistical power of population and investigative DNA profile databases^{18,19}. Although microsatellites and other simple tandem repetitive loci generate PCR-amplifiable alleles that can in principle be sized with precision on DNA sequencing gels, restricted allelic variability and the frequent occurrence of PCR-artefact bands limit their usefulness^{8,9,11}. There has also been considerable debate over the evaluation of the statistical weight of DNA profile evidence (see for example refs 18-20), in particular over the estimation of appropriate 'allele' frequencies from limited population databases and the assumption, required for genotype frequency estimation, that alleles associate at random in human populations^{16,19-21}.

We show here that many of these limitations of current DNA typing systems can, in principle, be overcome by assaying sequence variation in minisatellite alleles, rather than length differences.

Minisatellite variant repeat mapping

Minisatellite alleles frequently vary not only in repeat copy number but also in the interspersion pattern of variant repeat units along alleles (Fig. 1a)^{2,3,22-27}. Previous analysis of the hypervariable locus *DIS8* (probe MS32) showed two classes of repeat unit (designated a-type, t-type) that differ by a single base substitution which creates or destroys a *HaeIII* restriction site²⁸. Interspersion patterns of *HaeIII*⁺ and *HaeIII*⁻ repeat

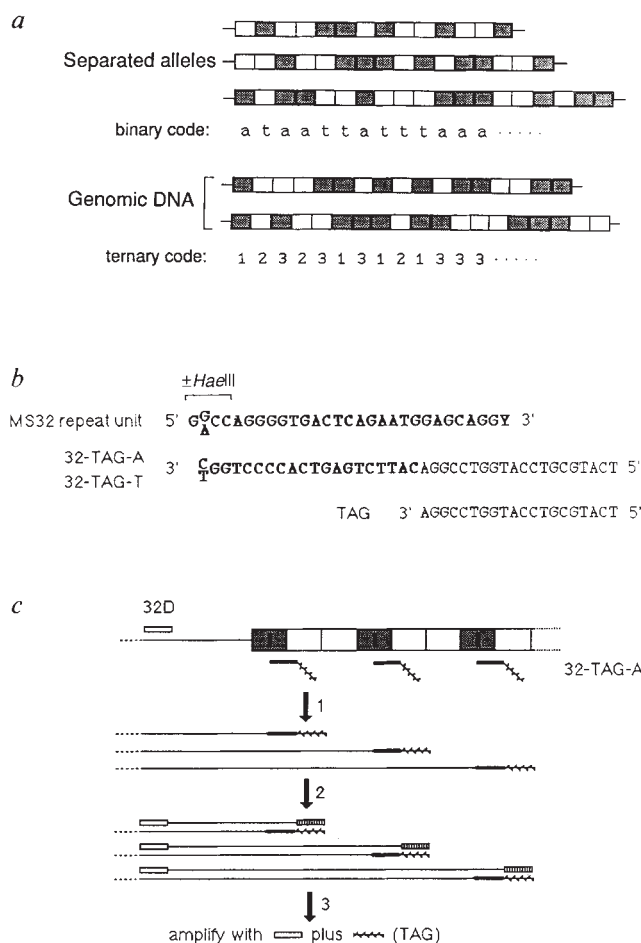


FIG. 1 The principles of minisatellite repeat coding. *a*, Minisatellite alleles consisting of interspersed arrays of two variant repeat units termed a-type (shaded boxes) and t-type (open boxes). Individual alleles can be encoded as a binary string extending from the first repeat unit. In total genomic DNA, a corresponding ternary code of both alleles superimposed can be generated. *b*, The consensus 29-bp repeat unit sequence of human minisatellite MS32 (*D1S8*) showing the polymorphic site that generates *HaeIII* cleavable (a-type) repeats and *HaeIII*-resistant (t-type) repeats. 32-TAG-A and 32-TAG-T are variant repeat specific oligonucleotides terminating at this polymorphic site. Each consists of 20-nt minisatellite repeat sequence (bold) preceded by a 20-nt 5' synthetic non-minisatellite extension identical to the TAG amplicon. *c*, The principle of MVR-PCR, illustrated for a single allele amplified using primer 32-TAG-A. (1) At low concentration of primer, 32-TAG-A will anneal to about one a-type repeat unit per target minisatellite molecule and extend into the flanking DNA. (2) Amplicon 32D primes from the flanking DNA, creating a sequence with an end (hatched box) complementary to TAG. (3) DNA fragments terminating in 32D and the TAG complement now amplify efficiently using high concentration of 32D and TAG amplicons, to create a stable set of PCR products extending from the flanking 32D site to each a-type repeat unit. Occasional internal priming off PCR products by 32-TAG-A will generate authentic but shorter PCR products. Use of primer 32-TAG-T at stage 1 will create a complementary set of products terminating at each t-type repeat unit.

We have now developed a much simpler MVR mapping system applicable to MS32 alleles of any length. The approach outlined in Fig. 1*b,c* uses two different MVR-specific primers which prime off either a-type or t-type repeat units. Amplification using one or other primer together with amplimer 32D from a fixed site in the minisatellite flanking DNA will generate two complementary sets of products from the ultravariabile end of any

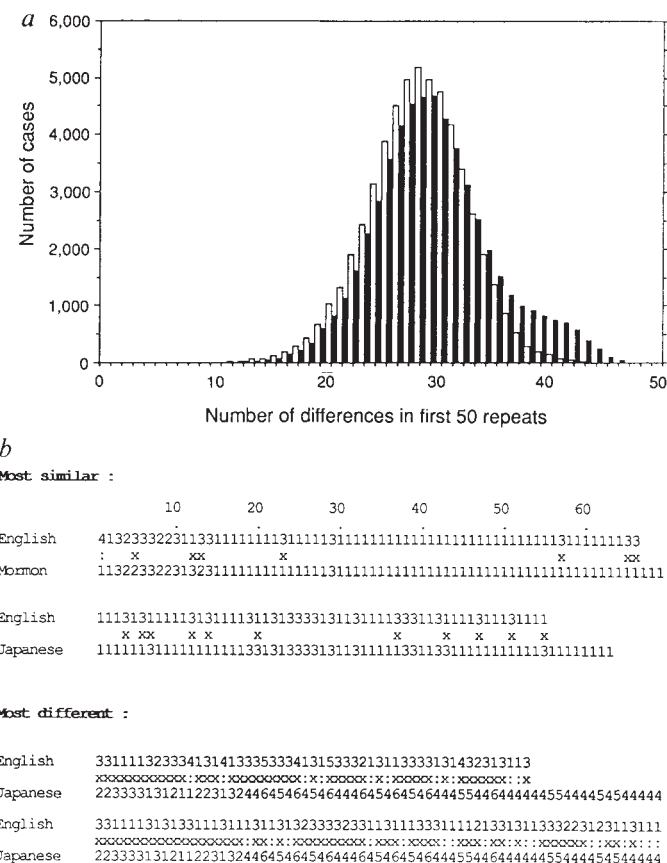


FIG. 3 Individual variation in diploid MVR codes. Codes extending for at least 50 repeat units were determined for 251 unrelated Caucasians (177 British, 20 French, 48 Utah Mormons, 2 Amish, 4 Venezuelan) and 83 Japanese. *a*, Filled bars, number of code differences seen over the first 50 repeat units, determined for every pairwise comparison of individuals (55,611 pairwise comparisons in total). Open bars, number of differences after removal of band intensity information, scoring code 1(aa) and 4(aO) as equivalent and code 2(tt) and 5(tO) as indistinguishable. Mean numbers of differences per pair of individuals were 30.1 and 27.9, respectively. *b*, Examples of the most similar and most dissimilar pairs of individuals over the first 50 repeat units, with discordances marked with x, or : for discordances which rely solely on intensity differences. The complete code scored per individual is shown, together with additional differences beyond repeat unit 50.

METHODS. All MVR codes were determined as described in Fig. 2. 62 ± 6 repeat units were scored per individual. If there was any doubt about the coding state at any given repeat position in an individual, the position was marked with a question mark. Only 0.3% of code positions (59 in 20,720 scored) were marked ? and were ignored in database searches. Data were stored as ASCII files and analysed using software written in VAX BASIC V3.4 and run on a VAX 8650 computer operating on VMS 5.3-1.

Allele

1 2 3

A T A T A T

Individual

4 5 6 7 8 9

A T A T A T A T A T

code position

bp

60 2004

50 1714

40 1424

30 1134

20 844

10 554

1 293

★

→

FIG. 2 Examples of minisatellite allele repeat coding by MVR-PCR on separated alleles (lanes 1–3) and total genomic DNA (lanes 4–9). DNA samples were amplified using 32-TAG-A (A) or 32-TAG-T (T) in a high concentration of primers 32D and TAG. PCR products were separated by agarose gel electrophoresis and detected by Southern blot hybridization. The first repeat unit (asterisk) is weakly detected and cannot be scored reliably. Diploid scoring on total genomic DNA therefore commences at the second repeat (code position 1). This start position for reading the code was confirmed by running a standard individual of known code on all gels. Null or O-type repeat units in allele 3 which do not amplify with either MVR-specific primer are arrowed. Individual 6 also shows an O-type repeat in one of his alleles, generating a code 4 (aO) position (star) detected as a relatively faint A track band with no band in the T track. This individual also contains a short allele of 28 repeat units, as shown by loss of codes 1, 2 and in particular 3 above the arrowed position and the presence of only code 4 (aO, faint A) and code 5 (tO, faint T) repeats, similar to the separated allele profiles. The presence of this short allele was confirmed by conventional Southern blot hybridization analysis of genomic DNA with MS32 (data not shown).

METHODS. Samples (100 ng) of genomic DNA, or equivalent amounts of 5–19-kb-long alleles separated from *Mbol* digests of genomic DNA by preparative gel electrophoresis, were amplified in 7- μ l reactions in 1 μ M primer 32D, 1 μ M primer TAG and either 10 nM 32-TAG-A or 20 nM 32-TAG-T, plus 0.25 unit AmpliTaq (Perkin–Elmer–Cetus), using the PCR buffer system described previously²⁸. Reactions were cycled for 1.3 min at 96°C, 1 min at 68°C, and 5 min at 70°C for 18 cycles on a DNA Thermal Cycler (Perkin–Elmer–Cetus), followed by a chase for 1 min at 67°C, 10 min at 70°C for two cycles. The sequence of the flanking primer 32D is 5'-CGACTCGCAGATGGAGCAATGGCC-3'²⁸. PCR products were electrophoresed through a 35 cm long 1% agarose (Sigma type I) gel in 89 mM Tris-borate (pH 8.3), 2 mM EDTA, 0.5 μ g ml⁻¹ ethidium bromide alongside ϕ X174 DNA $\times$ *Hae*III until the 118-bp marker had reached the end of the gel. DNA was denatured, transferred by blotting onto Hybond-N (Amersham) and hybridized to ³²P-labelled MS32 minisatellite probe for 3 h at 65°C as described previously². Autoradiography was for 6 h at room temperature.

MS32 allele, from which the MVR map can be deduced. To prevent progressive shortening at each PCR cycle because of MVR-specific primers priming internally in PCR products, MVR detection and subsequent amplification were uncoupled by providing each MVR-specific primer with a 20-nucleotide (nt) 5' extension 'TAG' and carrying out amplifications with a low concentration of one or other tagged primer and high concentrations of 32D and the TAG sequence itself.

Application of MVR-PCR at limited cycle number to MS32 alleles separated from genomic DNA generated continuous complementary ladders of PCR products detectable by Southern blot hybridization and extending >3 kb (100 repeat units) into each allele, from which allele binary codes could be read (Fig. 2). These were consistent with codes determined by partial digestion with *HaeIII* (data not shown). With additional PCR cycles, binary codes could be determined from PCR products directly visualized on ethidium bromide-stained gels, although overamplification and collapse of minisatellite PCR products⁶ limited coding to roughly the first 15 repeat units (data not shown).

Occasionally, a 'rung' on the MVR coding ladder failed to be amplified by either MVR-specific primer (Fig. 2), indicating the presence of 'null' repeats containing additional sequence variant(s) 3' to the *HaeIII* site which block priming by either primer. It was found that 1.6% of repeat units scored from 32 separated Caucasian alleles were null or O-type repeats, compared with 72.9% a-type repeats and 25.5% t-type repeats. O-type repeats tend to cluster in a limited number of alleles (see for example Figs 4 and 5), and can correspond both to *HaeIII*-cleavable and *HaeIII*-resistant repeat units.

MVR-PCR on genomic DNA

MVR-PCR on total genomic DNA produces a profile of both alleles superimposed, to generate a ternary code for two-variant alleles (Fig. 1a), where each rung in the ladder can be coded as 1(both alleles a-type at that position, aa), 2(both t-type, tt) or 3(heterozygous, at). The presence of O-type repeats creates three additional coding states, namely 4(aO), 5(tO) and 6(OO). The last will appear as a gap on the ladder. Coding states 4-6 will also be generated beyond the end of the shorter allele, as the code will be derived from only one allele. No PCR products

will appear beyond the end of the longer allele, generating a 66666... code.

Examples of MVR-PCR on total genomic DNA are shown in Fig. 2. In each case, diploid codes could be read at least 50 repeat units into the minisatellite. The two tracks generating the code contain considerable informational redundancy; in almost all cases, an intense band in the A track was matched by no band in the T track (code 1, aa), a faint A band by a faint T band (code 3, at) and no A band by an intense T band (code 2, tt). This dosage phenomenon provides a detailed check on the authenticity of the code generated, and also makes it possible to identify with good reliability individuals with short alleles and rung positions which are heterozygous for a null or O-type repeat (code 4, aO; code 5, tO) (Fig. 2).

Variation in diploid codes

MVR-PCR typing of 334 unrelated individuals showed that there were on average 30 code mismatches per pair of individuals over the first 50 repeat units (Fig. 3a). No two individuals shared the same MVR code, and all individuals could be distinguished using only the first 17 repeat positions. Individual specificity remained when band intensity information was removed by converting all code 4(aO) and code 5(tO) positions to codes 1(aa) and 2(tt), respectively, to generate quaternary codes (1, 2, 3, 6) corresponding to bands present only in the A track, only in the T track, in both tracks and in neither track, respectively. The two most similar individuals had MVR codes dominated by code 1(aa) (Fig. 3b), indicating that all four alleles in these two individuals were composed largely of a-type repeats; such homogeneous alleles have been noted previously²⁸. The most dissimilar pairs of individuals arose where one contained a short allele, creating a diploid code dominated by the rare codes 4, 5 and 6. Short (<50 repeats) alleles with allele lengths ranging from 19 to 44 repeat units were found in 7.8% of individuals. Short alleles do not occur with equal frequency in all populations; thus 5.6% of Caucasians typed contained short alleles, compared with 23% of Japanese ($P < 0.001$).

Heterozygosity levels at MS32

Diploid codes provide an objective method for identifying homozygotes, in contrast to allele length analysis of minisatel-

FIG. 4 Reconstruction of the MVR codes of individual MS32 alleles by pedigree analysis. *a*, Incomplete reconstruction of the haplotypes of all four parental alleles using a single child. The paternal and maternal alleles transmitted to the child are labelled A and C, and the nontransmitted alleles B and D. Parental haplotypes can be deduced, assuming no recombination between parental alleles, at all positions except where the father, mother and child are heterozygous for the same repeat unit types. Such ambiguities are indicated by a question mark. *b*, Complete reconstruction of allele maps using the large sibship of CEPH family 104. The children show four different diploid codes, corresponding to the four possible combinations of parental alleles. The mother unusually contains two short alleles, resulting in coding state 6(OO) beyond the end of the longer allele. The resulting haplotypes (alleles C and D) terminate in a string of null (nonexistent) repeats. Both of these alleles have also been mapped by partial digestion with *HaeIII*²⁸ (in bold).

METHODS. Haplotypes were extracted using software written in VAX BASIC V3.4. For a family with a single child, the four parental haplotypes were extracted sequentially along each position of the diploid code. For each position, the codes of the father, mother and child were checked in a look-up table to determine whether exclusions exist and, if not, to determine the repeat unit types transmitted or not from each parent to the child. For example, if the father is 1(aa), mother 3(at) and child 3(at), then no exclusions exist and the repeat unit at that position on each

Individual	Alleles	MVR code position									
		5	10	15	20	25	30	35	40	45	50
a Single child											
father	A B	11111	13131	11133	12131	31333	33311	13111	11131	11113	11113...
mother	C D	11113	11111	31123	14331	41311	41334	13143	31133	11311	11111...
child	A C	11113	11111	31123	13331	31233	33131	13113	31123	11313	11113...
	allele A	?aaaaa	aaaaa	aaat?	ata?a	tattt	ttaaa	a?aaa	aaata	aaaat	aaaat...
	allele B	?aaaaa	atata	aaaa?	ata?a	aaaaa	ataaa	a?aaa	aaaaa	aaaaa	aaaaa...
	allele C	?aaaat	aaaaa	taat?	aat?a	ataaa	aaaaa	a?aat	taatt	ataaa	aaaaa...
	allele D	?aaaaa	aaaaa	aaat?	a0a?a	0aaaa	0ata0	a?a0a	aaaaa	aaaaa	aaaaa...
b Multiple children											
father 10401	A B	33331	33131	33123	33111	11332	11333	31113	23131	22331	33133...
mother 10402	C D	11233	23251	14331	15112	35555	55454	55555	55666	66666	66666...
children:											
10406, 07, 11, 12	A C	11313	31331	31321	13113	34445	44544	44444	54454	55554	44454...
10408	A D	11331	33341	34131	14113	13332	33331	33333	23454	55554	44454...
10404, 05	B C	33233	23321	13323	32113	34555	44455	54445	55444	55444	55445...
10409, 10	B D	33221	22351	15133	35113	13222	33123	23332	22444	55444	55445...
	allele A	?aaaaa	aaaaa	taata	aaaaa	aaaat	aataa	aaaaa	taata	tttta	aaata...
	allele B	?tttta	ttata	atatt	ttaaa	aattt	aaatt	taaat	ttaaa	ttaaa	ttaat...
	allele C	?aatat	tatta	aatta	ataat	t0000	00000	00000	00000	00000	00000...
allele C <i>HaeIII</i> map		aaatat	tatta	aatta	ataat	t					
allele D		?aatta	ttt0a	a0aaa	a0aat	atttt	ttata	ttttt	tt000	00000	00000...
allele D <i>HaeIII</i> map		aaatta	tttaa	aaaaa	aaaat	atttt	ttata	ttttt	tt		

allele is given by allele A, a; allele B, a; allele C, t; allele D, a. By contrast, codes 1(aa)+2(tt)+5(tO) would give a paternal mutation/exclusion. For families with more than one offspring, the incomplete haplotypes of each parental allele were determined separately for each child and then melded into a complete unambiguous haplotype.

preceded and followed by MVR code derived from the same maternal allele. The mutation event is therefore probably intra-allelic, and could have arisen for example by unequal sister chromatid exchange or replication slippage. Mutants a, b and g also seem to have arisen by an intra-allelic event, but the presumptive exchange point lies too close to the beginning of the allele to be certain that the mutant allele does not contain a recombinant haplotype. By contrast, mutant e (Fig. 6a-c) provides clear evidence for interallelic unequal exchange

between the two paternal alleles, the mutant allele commencing with one paternal haplotype, then switching after two a-type repeats of unknown origin to the beginning of the other paternal allele. Mutant f also seems to have arisen by interallelic unequal exchange. DNA markers flanking MS32 are insufficiently close and informative to test whether these two mutations have been accompanied by exchange of distal flanking markers²⁹.

A recombination hot-spot?

Previous studies of minisatellite mutation have shown that allele length change events are largely, if not completely, restricted to single alleles²⁸⁻³⁰. Our data provide the first direct evidence that interallelic recombination or gene conversion can have a major role in minisatellite instability. If the two apparently interallelic mutation events have arisen by a conventional (if unequal) interallelic recombination process, this implies a recombination frequency of $2/572 = 0.3$ centiMorgans (cM) over the first ~400 base pairs (bp) of the minisatellite, a 700-fold enhancement over the mean frequency of 1cM per 10⁶ bp in the human genome. If correct, this would represent a dramatic example of a human recombination hot-spot, and revitalizes earlier speculation that minisatellites may be actively involved in chromosomal processes such as homologue recognition, synapsis and meiotic recombination^{3,31,32}.

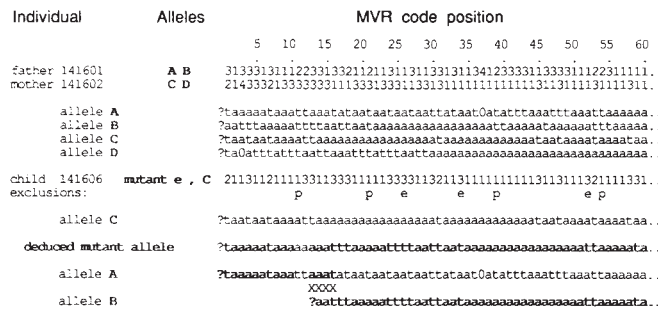
Forensic considerations

MVR-PCR is simple, and preliminary investigations suggest that it should be applicable to forensic analysis. It is sensitive, generating normal profiles down to 10 ng human DNA. Below this level, apparently random fluctuations in band intensity can arise, owing to stochastic loss of PCR products from the small number of input molecules; reliable consensus diploid codes can nevertheless be obtained by comparing replicate MVR profiles from subnanogram amounts of genomic DNA. MVR-PCR can also be applied to partially degraded DNA, as it recovers information from any DNA fragments long enough to include the flanking 32D priming site plus at least some 29 bp repeat units. Additional information can be recovered by substituting the flanking primer 32D for 32O (sequence GAG-TAGTTTGGTGGGAAGGGTGGT), which primes immediately adjacent to the start of the tandem-repeat array (Fig. 1c). Although degraded DNA yields a truncated diploid code, it is still compatible with database searches though with reduced discriminating power.

MVR-PCR can also be applied to mixed DNA samples (for example victim plus rapist DNA recovered from semen-bearing vaginal swabs), particularly if pure DNA from one of the two individuals (for example victim) is available. DNA mixing experiments have shown that 10% admixture can be detected, and that comparison of 'victim' and mixed DNA samples can yield an ambiguous diploid code of the 'rapist' (Fig. 7). For example, if the victim is code 1 (aa) and the mixture contains an additional band in the T track, then the rapist must be code 2 (tt), 3 (at) or 5 (tO). The efficiency of identification using mixed DNA information was assessed by creating 2 × 10⁶ different combinations of 'victim', 'rapist' and 'false suspect' from the database of MVR codes. For each case, the ambiguous rapist code deducible from a victim-rapist mixture was checked against the suspect for exclusions. On average, 14 exclusions per case were detected over the first 50 repeat units, and only 14 out of 2 × 10⁶ 'false suspects' failed to show any exclusion (99.9993% mean exclusion rate).

MVR-PCR could also be used in parentage testing because the diploid code of a nonparent will frequently show exclusionary mismatches with the child (see for example Fig. 6a). To determine the effectiveness of MVR-PCR in excluding non-fathers, 28,635 Caucasian mother-child-nonfather trios were created from the MVR code database and analysed for paternal exclusions. On average, 9.9 exclusions were obtained over the first 50 repeats, of which 4.7 were paternal-specific and the

a Mutation mapping in pedigrees



b Nature of mutations

mutation	CEPH individual	change in repeat unit copy no.	detected on Southern blot	mechanism
maternal				
a	134505	+1	-	intra-allelic?
b	134606	+1	-	intra-allelic?
c	140808	+?	+	?
d	142106	+1	-	intra-allelic
paternal				
e	141606	+13	+	inter-allelic
f	142409	+3	+	inter-allelic?
g	1329405	+2	+	?

mutation rate = 7/572 per gamete = 0.0122 (95% confidence limits 0.006-0.023)

c Location of presumptive exchange points

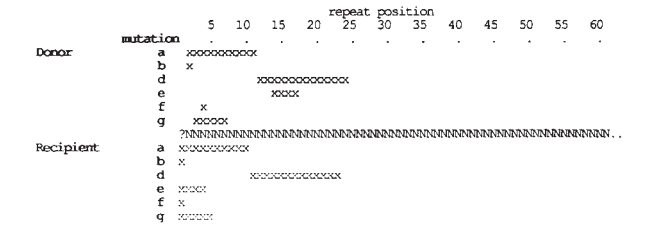


FIG. 6 Analysis of MS32 mutant alleles detectable in pedigrees. a, A CEPH pedigree showing a child with a mutant allele. Maps of parental alleles A–D were deduced from seven nonmutant offspring (not shown). Comparison of the diploid code of child 141606 with the parents shows four specifically paternal exclusions (p) plus three ambiguous exclusions (e) which do not indicate the parental origin of the mutant allele. There are no maternal exclusions, and thus the child has inherited a mutant paternal allele and nonmutant maternal allele. The diploid code of the child is compatible with the child having inherited maternal allele C but not D. Subtraction of the code for allele C from the diploid code of the child yields the code for the mutant paternal allele. Comparison with paternal alleles A and B indicates that the mutant allele commences with the code of allele A and then switches to the beginning of the code of allele B after two a-type repeats of unknown origin. This allele therefore seems to have arisen by unequal crossing over between the two paternal alleles, as indicated, with possible cross-over sites marked X. b, Summary of mutant alleles detected in 286 offspring from the CEPH panel of families. This survey detected all allele length change mutations previously detected by Southern blot analysis of *AluI* digests of genomic DNA³⁴, plus three previously undetected mutations resulting from gains of a single repeat unit. In all cases, the change in repeat unit copy number relative to the recipient allele (the allele contributing to the end of the mutant allele) is consistent with allele length changes detected by Southern blot analysis (not shown). Mutant c is complex and has yet to be fully characterized. c, Possible locations of unequal exchange points on the donor allele (the allele which contributes to the beginning of the mutant allele) and the recipient allele, as shown for mutant e in (a) above. For presumptive intra-allelic (sister chromatid) unequal exchange, the donor and recipient alleles are identical.



FIG. 7 Incomplete MVR code information recoverable from mixed DNA samples. Two individuals (X, Y) were chosen at random from a collection of 450 typed people, and their identities were concealed from the analyst. Samples (100 ng) of genomic DNA from X and Y, or from X plus Y mixed in the indicated proportions, were amplified by MVR-PCR for 18 cycles and PCR products detected by Southern blot hybridization. S, Standard individual included on all gels. By comparing the MVR-PCR profile of X with the profile of the mixed DNA samples, possible genotypes of Y can be deduced, as indicated, at all repeat positions where X is not code 3(at), by checking for A- or T-track specific bands present in the mixture but not X. For example, if the mixture but not X contain a band in the T-track at a given position, then the possible codes of Y at that position are 2(tt), 3(at) or 5(tO). The MVR code of X and the incomplete and ambiguous MVR code of Y deduced from the mixed DNA samples were screened across the database of 450 individuals to reveal, correctly and uniquely, the identities of X and Y.

between a forensic sample and a criminal suspect, and for creating very large communal population and investigative DNA databases. As there are far in excess of 3,500 different alleles ($\gg 6 \times 10^6$ diploid codes), it is likely that very large databases can be constructed before any significant saturation of MVR code types occurs. (Note that this does not hold for pairs of siblings who will have a roughly 1/4 chance of sharing the same parental alleles and therefore MVR code.) Very large databases provide a simple method for determining the statistical significance of a match between a forensic sample and a suspect, by counting the frequency (probably zero) of the particular MVR code in the appropriate database. This approach uses observed phenotype frequencies, rather than genotype frequencies deduced from limited population databases under assumptions of Hardy-Weinberg equilibrium.

MVR-PCR can be used on any minisatellite locus showing internal variation, provided that variant repeats with abnormal repeat length do not exist, as seems to be the case with MS32 but not with most other minisatellites^{2,22-25,33}. Abnormal length repeats will throw the MVR ladders of the two constituent alleles out of register, making at least part of the diploid coding ladder uninterpretable. We have identified a few minisatellites which may be suitable for MVR-PCR and ultimately for multiplex MVR-PCR where several loci are amplified simultaneously to improve specificity, particularly in parentage analysis.

Discussion

MVR-PCR provides a novel and simple method for generating unambiguous and highly discriminatory digital information from human DNA. Many of the problems associated with other DNA typing approaches, such as the definition of DNA profile matching and measurement errors in allele sizing, are obviated. MVR-PCR has also provided a clear view of the extraordinary levels of allelic variability that can exist in human minisatellites, and revealed that recombination is involved in the generation of ultravariability. □

remainder directionally ambiguous. At least one paternal-specific exclusion was shown by 98.9% of nonfathers, and 99.8% showed at least one exclusion in total (paternal-specific plus ambiguous). The efficiency of paternity testing is, however, limited by the high *de novo* mutation rate at MS32, and by the need to correctly identify heterozygous null positions in the child (code 4, aO; 5, tO) from interpretation of band intensities.

Diploid MVR coding potentially offers many advantages over currently used DNA typing systems that involve allele length measurements. MVR profiles contain considerable informational redundancy enabling code authenticity to be checked. Code generation does not require standardization of electrophoretic systems, is immune to gel distortions and band shifts, does not involve error-prone DNA fragment length measurement, and does not require side-by-side comparisons of DNA samples on the same gel. MVR-PCR generates digital DNA typing information ideal for determining objectively a match

Received 29 August; accepted 23 October 1991.

1. Nakamura, Y. *et al Science* **235**, 1616-1622 (1987).
2. Wong, Z., Wilson, V., Patel, I., Povey, S. & Jeffreys, A. J. *Ann. hum. Genet.* **51**, 269-288 (1987).
3. Jeffreys, A. J., Wilson, V. & Thein, S. L. *Nature* **314**, 67-73 (1985).
4. Jeffreys, A. J., Wilson, V. & Thein, S. L. *Nature* **316**, 76-79 (1985).
5. Jeffreys, A. J., Turner, M. & Debenham, P. *Am. J. hum. Genet.* **48**, 824-840 (1991).
6. Jeffreys, A. J., Wilson, V., Neumann, R. & Keyte, J. *Nucleic Acids Res.* **16**, 10953-10971 (1988).
7. Boerwinkle, E., Xiong, W., Fourest, E. & Chan, L. *Proc. natn. Acad. Sci. U.S.A.* **86**, 212-216 (1989).
8. Litt, M. & Luty, J. A. *Am. J. hum. Genet.* **44**, 397-401 (1989).
9. Tautz, D. *Nucleic Acids Res.* **17**, 6463-6471 (1989).
10. Weber, J. L. & May, P. E. *Am. J. hum. Genet.* **44**, 388-396 (1989).
11. Hagelberg, E., Gray, I. C. & Jeffreys, A. J. *Nature* **352**, 427-429 (1991).
12. Jeffreys, A. J., Brookfield, J. F. Y. & Semeonoff, R. *Nature* **317**, 818-819 (1985).
13. Jeffreys, A. J. *et al. in DNA Fingerprinting: Approaches and Applications* (eds Burke, T., Dolf, G., Jeffreys, A. J. & Wolff, R.) 1-19 (Birkhäuser, Basel, 1991).
14. Baird, M. *et al. Am. J. hum. Genet.* **39**, 489-501 (1986).
15. Balazs, I., Baird, M., Clyne, M. & Meade, E. *Am. J. hum. Genet.* **44**, 182-190 (1989).
16. Odelberg, S. J. *et al. Genomics* **5**, 915-924 (1989).
17. Smith, J. C. *et al. J. forensic Sci. Soc.* **30**, 19-32 (1990).
18. Lander, E. S. *Nature* **339**, 501-505 (1989).
19. Budowle, B. *et al. Am. J. hum. Genet.* **48**, 841-855 (1991).
20. Devlin, B., Risch, N. & Roeder, K. *Science* **249**, 1416-1420 (1990).

21. Chakraborty, R., Fornage, M., Gueguen, R. & Boerwinkle, E. *in DNA Fingerprinting: Approaches and Applications* (eds Burke, T., Dolf, G., Jeffreys, A. J. & Wolff, R.) 127-143 (Birkhäuser, Basel, 1991).
22. Owerbach, D. & Aagaard, L. *Gene* **32**, 475-479 (1984).
23. Jarman, A., Nicholls, R. D., Weatherall, D. J., Clegg, J. B. & Higgs, D. R. *EMBO J.* **5**, 1857-1863 (1986).
24. Wong, Z., Wilson, V., Jeffreys, A. J. & Thein, S. L. *Nucleic Acids Res.* **14**, 4605-4616 (1986).
25. Nakamura, Y. *et al. Nucleic Acids Res.* **15**, 2537-2547 (1987).
26. Page, D. C. *et al. Genomics* **1**, 243-256 (1987).
27. Gray, I. C. & Jeffreys, A. J. *Proc. R. Soc. B* **243**, 241-253 (1991).
28. Jeffreys, A. J., Neumann, R. & Wilson, V. *Cell* **60**, 473-485 (1990).
29. Wolff, R. K., Plaetke, R., Jeffreys, A. J. & White, R. *Genomics* **5**, 382-384 (1989).
30. Wolff, R., Nakamura, Y. & White, R. *Genomics* **3**, 347-351 (1988).
31. Chandley, A. C. & Mitchell, A. R. *Cytogenet. Cell Genet.* **48**, 152-155 (1988).
32. Royle, N. J., Clarkson, R. E., Wong, Z. & Jeffreys, A. J. *Genomics* **3**, 352-360 (1988).
33. Simmler, M. C. *et al. EMBO J.* **6**, 963-969 (1987).
34. Armour, J. A. L., Patel, I., Thein, S. L., Fey, M. F. & Jeffreys, A. J. *Genomics* **4**, 328-334 (1989).

ACKNOWLEDGEMENTS. We thank J. Armour, A. Collick, N. Royle, I. Gray and R. Dalgleish for helpful discussions, D. Langton for oligonucleotide synthesis and H. Cann, J. Dausset, Y. Katsumata and P. Debenham for providing human DNA samples. This work was supported by grants to A.J.J. from the Medical Research Council and Wolfson Foundation. A.J.J. is a Lister Institute Research Fellow. This work is the subject of Patent applications.

Origin of kinematic subsystems in elliptical galaxies

Lars Hernquist* & Joshua E. Barnes†

* Lick Observatory, University of California at Santa Cruz, Santa Cruz, California 95064, USA

† Institute for Astronomy, University of Hawaii, Honolulu, Hawaii 96822, USA

ELLIPTICAL galaxies were once thought to be smooth, featureless stellar systems with little or no substructure, but increasingly sophisticated observations are challenging this point of view. Some ellipticals contain small central disks which may be counter-rotating or otherwise kinematically decoupled from the rest of the galaxy^{1–6}. It seems unlikely that these disks could form during the monolithic collapse of a slowly rotating proto-galaxy, as the most plausible outcome of such evolution is a system with a simple rotation pattern. Like other kinds of fine structure in elliptical galaxies⁷, these subsystems have been widely interpreted as evidence for multiple formation events or episodes. Here we demonstrate the formation of a counter-rotating central gas disk in a merger of two gas-rich disk galaxies of equal mass. Such a structure may well account for the unusual gas kinematics found in the merger remnant NGC7252 (refs 8, 9). Continued star formation in such gaseous disks may produce central components with decoupled kinematics, resembling the cores of some elliptical galaxies.

Most of the light emitted by an elliptical galaxy originates in a smooth component with ellipsoidal contours and a surface density profile well approximated by the $R^{1/4}$ law of de Vaucouleurs¹⁰. Recent surveys indicate, however, that most ellipticals also possess weak features, including 'shells,' boxy isophotes, X-structures, plumes, and other kinds of fine structures⁷. In some cases these structures are distinct from the smooth light profile, suggesting that they form by a secondary process involving an accretion or merger, a concept supported by observational and theoretical considerations^{9,11,12}. If this view is correct then environmental effects play an important role in structuring present-day galaxies.

A related phenomenon, discovered by studying the central regions of nearby ellipticals, is that some contain cores displaying velocity fields different from those of the surrounding galaxies proper. Most bright ellipticals rotate slowly and are supported mainly by an anisotropic velocity dispersion^{13–15}. Yet ellipticals such as NGC5813 contain rapidly rotating cores¹, whereas the central regions of others such as IC1459 and NGC5322 rotate in an opposite sense to the weak but observable rotation of the galaxies as a whole^{2,3}. In still other examples, the rotation axis of the core is nearly perpendicular to that of the galaxy^{4–6}. Statistics of the frequency of such unusual velocity fields remain ambiguous, but surveys imply that at least 20–25% of all ellipticals have these structures^{2,16,17}.

It is natural to suppose that such kinematic features resulted from merger or accretion events. This supposition may not always be correct, as triaxial galaxies with substantial streaming can seem to contain counter-rotating cores as a result of projection effects¹⁸. But in cases such as IC1459 (ref. 2), NGC4621 (ref. 19), 5322 and 3610 (ref. 20), line-profile analysis indicates that the velocity fields observed are due to the presence of a central component with kinematics unlike that of the main bodies of these galaxies. Kormendy²¹ suggested that the central rotation of NGC5813 might have resulted from the accretion of a small dwarf galaxy which sank to the centre of the primary by dynamical friction and that the rotation reflects the spin of the dwarf. Numerical simulations²² support this conjecture but indicate that the initial spin of the dwarf is largely irrelevant and that the rotation of the core is essentially a residue of the dwarf's orbital angular momentum.

Here, we present a numerical model illustrating another possible origin for these cores—mergers between gas-rich disk galaxies of comparable mass. As the models presented here do not include the effects of star formation, we cannot address the structure of stellar cores in detail. Nonetheless, our results indicate that such mergers can produce remnants that generally resemble elliptical galaxies and contain central concentrations of gas kinematically decoupled from the stellar component. Stars forming from this gas would very probably be distributed in disks, as suggested by observations.

This model is one of a series run using a code²³ combining N -body^{24,25} and smoothed particle hydrodynamic^{26,27} methods to simulate self-gravitating systems of gas, stars and dark matter. The interstellar medium is treated as a single-phase fluid, with radiative cooling cut off below 10^4 K to inhibit a local Jeans' instability. Shocks, numerically broadened by a standard form of artificial viscosity, provide the main source of heat input to the gas, which quickly cools back to the cut-off temperature.

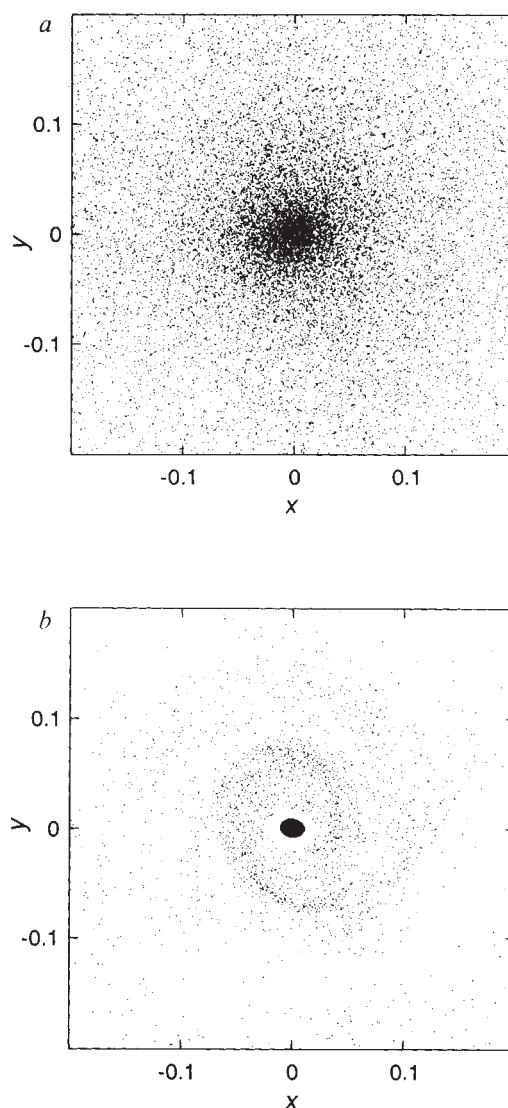


FIG. 1 Distribution of stars (a) and gas (b) in inner regions of remnant resulting from a disk-disk merger. The gas comprises two nested disks. The view is from a direction perpendicular to the outer diffuse disk of gas. Results are displayed in a dimensionless system of units in which the radial exponential scale-length of the progenitor disk galaxies is $1/12$. Scaled to the properties of the Milky Way, each panel measures 16.8 kpc per edge, and the inner gas disk is confined to a region ~ 800 pc across, whereas the outer gas disk is ~ 5.5 kpc in diameter.

The initial conditions and numerical parameters for this run are similar to those used previously²⁸, with the following changes. First, the two disk galaxies were given antiparallel spin vectors, tilted by 71° from the orbital axis and 19° from the separation direction at first pericentre. Second, the pericentric separation for the ingoing parabolic orbit was increased from 0.2 to 0.4 (in units²⁹ such that the half-mass radii of the original galaxies are ~ 0.25). Third, because such wider encounters take longer to merge, the total particle number N was reduced from 90,112 to 45,056, with the fraction devoted to each component held constant; despite this, two-body relaxation is not significant over the relatively short timespan of this calculation. Otherwise, as in other runs, the two galaxies employed are composite bulge/disk/halo models²⁹ with mass ratios of bulge: disk: halo = 1:3:16, and 10% of their disk mass in gas.

The smaller number of particles run enabled us to continue this calculation for roughly one time unit past the eventual merger of the system (comparable to one rotation period at the half-mass radii of the original galaxies). By the end of this time this merger remnant has settled into an approximate equilibrium within a radius of ~ 0.5 length units, but is still accreting material from larger radii. Figure 1 shows the distribution of stars and gas in the remnant at the end of the calculation. Particles representing the stellar component (bulge plus disk) show a good approximation to a de Vaucouleurs' law in projection; this remnant, like others^{28,29}, is photometrically similar to an elliptical galaxy. Axial ratios, measured from the luminous particles within the most tightly bound third of the remnant, are $\sim 0.8:0.9:1$; the minor axis of the remnant is tilted by $\sim 50^\circ$ with respect to the original orbital axis.

The particles representing the gas, shown in Fig. 1b, are distributed quite differently from the stars. Most of the gas is concentrated in two nested disks, the inner one being the blacked-out oval at the centre of the frame. This inner disk contains $\sim 60\%$ of all the gas in the system, or $\sim 7 \times 10^9 M_\odot$ if the original disks are roughly scaled to the Milky Way. The outer gas disk, which has an open spiral pattern in Fig. 1, has a somewhat larger scale height as shown in the edge-on view of Fig. 2a. The rotation axes of these inner and outer disks are tilted by $\sim 15^\circ$ and 4° , respectively, from the minor axis of the remnant itself. Further evolution of the system would probably bring the disks and the inner portion of the remnant into yet closer alignment with each other.

Of greatest interest to the present discussion are the rotation patterns of the two disks. Figure 2b shows gas particle velocities along the same line of sight used in Fig. 2a. It is evident that the massive inner gas disk counter-rotates with respect to the outer one. In view of the history of this remnant, such counter-rotation is not too surprising. Briefly, the inner disk is largely composed of the gas that collected at the centres of the original disks just after first passage^{28,30}. These central gas concentrations, which have masses and dimensions comparable to those observed in nuclear starburst galaxies, eventually collide nearly head-on when the galaxies finally merge. Thus this material has of necessity been subjected to strong gravitational torques in order to arrive at the centre of the remnant, and the extremely small amount of angular momentum it retains bears little or no relation to its initial orbital and rotational angular momentum. In contrast, the gas in the outer disk, like the stellar component of the remnant, has not been subjected to such strong torques and so retains some 'memory' of its original angular momentum. Thus the outlying gas rotates in the same general direction as the stellar remnant, but both are decoupled from the gas in the dense inner disk.

Although we have succeeded in producing a remnant resembling an elliptical galaxy with a kinematic subsystem, the correspondence between the simulations and actual galaxies remains somewhat uncertain; for one thing, we have ignored effects related to star formation. Counter-rotating cores like those in IC1459 consist mainly of stars. In fact, residual gas in the core

of IC1459 seems to rotate in the opposite direction from the stars there³. A similar distribution of gas and stars might result in the remnant shown in Figs 1 and 2a if we were simply to convert all the gas in the inner, counter-rotating disk into stars. But until we perform additional simulations which allow the gas to turn into stars and include feedback from supernovae, we can only speculate as to the final distribution of newly formed stellar material in the cores of merger remnants. Likewise, we cannot predict with any confidence how commonly counter-rotating cores arise during mergers of gas-rich disks. All the cases we have considered so far yield remnants with dense central cores of gas; at least 25% of these cores counter-rotate with respect to the remnant itself. We have examined only a small subset of parameter space, however, so this estimate is highly uncertain.

Nevertheless, it is of interest to note that the morphology of the remnant shown in Fig. 1 shares a number of qualitative features with IC1459, suggesting that this galaxy formed by a merger of two progenitors of comparable mass. In particular, IC1459 is a radio source, has a photometric twist, and contains faint wisps or arms of stellar debris at large radius³¹. It is believed that many radio galaxies are triggered by mergers³². Likewise, fine structures such as wisps seem to be a generic consequence of accretion events. In view of our numerical results, we suggest that the stellar core in IC1459 was produced by a nuclear starburst, provoked in turn by the merger that formed this galaxy.

At present, our model is probably more directly relevant to objects like NGC7252. This galaxy, although photometrically similar to an elliptical, features two long tails and is otherwise disturbed; these properties suggest⁸ that it was produced by a principal merger within the past $\sim 10^9$ years. The rotation curve of the ionized gas in the inner regions of NGC7252 is fairly complex and counter-rotates over a large radial interval⁸. Near its centre is a bright disk of ionized gas, and it also contains significant quantities of molecular gas with a total H_2 mass of $\sim 10^{10} M_\odot$ (ref. 9). An obvious interpretation is that NGC7252 is the aftermath of a merger between two gas-rich disks galaxies,

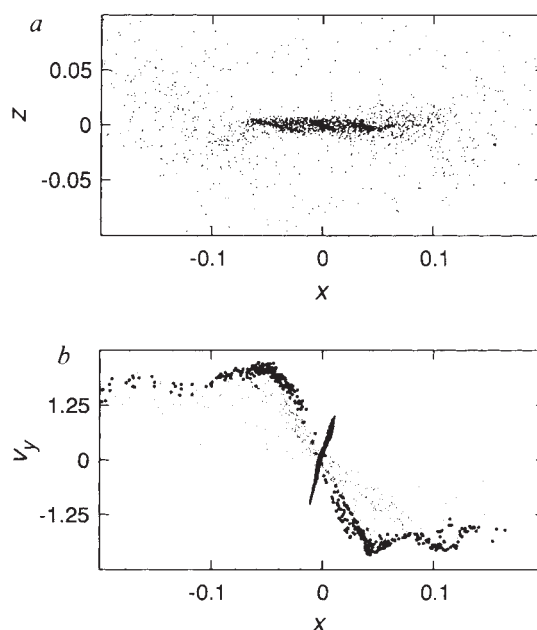


FIG. 2 a, Distribution of gas in remnant shown in Fig. 1, edge-on to the inner disk of gas. Results are displayed in the same system of units as in Fig. 1. b, Line-of-sight velocity through the gas shown in a. Particles having $|y| \leq 0.025$ are enlarged to highlight the rotation curve. Scaled to the properties of the Milky Way, the system of units is such that unit length and velocity correspond to 4.2 kpc and 160 km s^{-1} , respectively.

like the computer model described here, with a large fraction of the gas driven to the centre of the remnant, forming the disk seen there. The distribution and kinematics of the gas in NGC7252 are reminiscent of the gas shown in Figs 1 and 2. It therefore seems likely that the counter-rotation observed in NGC7252 is a signature of a major merger, rather than the accretion of a small companion as may be the case in NGC5813 (ref. 21). Although we cannot yet predict the long-term evolution of systems like that in Fig. 1 and 2, it is possible that star formation in the central gas disk of NGC7252 will eventually yield a counter-rotating stellar core much like that seen in IC1459.

If our interpretation is correct, then the discovery of kinematically distinct cores provides further evidence that many elliptical galaxies are merger products. This point of view is supported by observations implying that some ellipticals with such cores show independent signs of an interaction or a merger, including shells, radio activity and tidal tails^{8,9,33}. But not all ellipticals with peculiar core kinematics show signs of recent interactions; most, in fact, are otherwise rather ordinary. As gas dynamics is an essential aspect of the origin of kinematic subsystems in our computer simulations, our results support the notion that some normal ellipticals formed through mergers of gas-rich objects, perhaps like present-day spiral galaxies³⁴. □

Received 11 July; accepted 8 October 1991.

1. Efsthathiou, G., Ellis, R. S. & Carter, D. *Mon. Not. R. astr. Soc.* **201**, 975–990 (1982).
2. Bender, R. in *Dynamics and Interactions of Galaxies* (ed. Wielen, R.) 232–243 (Springer, Berlin, 1990).

3. Franx, M. & Illingworth, G. D. *Astrophys. J.* **327**, L55–60 (1988).
4. Franx, M., Illingworth, G. D. & Heckman, T. M. *Astrophys. J.* **344**, 613–636 (1989).
5. Jedrzejewski, R. I. & Schechter, P. L. *Astr. J.* **98**, 147–165 (1988).
6. Wagner, S. J., Bender, R. & Möllenhof, C. *Astr. Astrophys.* **195**, L5–8 (1988).
7. Seitzer, P. & Schweizer, F. in *Dynamics and Interactions of Galaxies* (ed. Wielen, R.) 270–271 (Springer, Berlin, 1990).
8. Schweizer, F. *Astrophys. J.* **252**, 455–460 (1982).
9. Schweizer, F. in *Dynamics and Interactions of Galaxies* (ed. Wielen, R.) 60–71 (Springer, Berlin, 1990).
10. de Vaucouleurs, G. *Ann. Astrophys.* **11**, 247–287 (1948).
11. Hernquist, L. & Quinn, P. J. *Astrophys. J.* **331**, 682–698 (1988).
12. Hernquist, L. & Quinn, P. J. *Astrophys. J.* **342**, 1–16 (1989).
13. Bertola, F. & Capaccioli, M. *Astrophys. J.* **200**, 439–445 (1975).
14. Illingworth, G. D. *Astrophys. J.* **218**, L43–47 (1977).
15. Binney, J. J. *Mon. Not. R. astr. Soc.* **183**, 501–514 (1978).
16. Illingworth, G. D. & Franx, M. in *Dynamics of Dense Stellar Systems* (ed. Merritt, D.) 13–23 (Cambridge University Press, 1989).
17. Jedrzejewski, R. I. & Schechter, P. L. in *Dynamics of Dense Stellar Systems* (ed. Merritt, D.) 25–30 (Cambridge University Press, 1989).
18. Statler, T. *Astr. J.* **102**, 882–892 (1991).
19. Bender, R. *Astr. Astrophys.* **229**, 441–451 (1990).
20. Rix, H.-W. & White, S. D. M. *Mon. Not. R. astr. Soc.* (in the press).
21. Kormendy, J. *Astrophys. J.* **287**, 577–585 (1984).
22. Balcells, M. & Quinn, P. J. *Astrophys. J.* **361**, 381–393 (1990).
23. Hernquist, L. & Katz, N. *Astrophys. J. Suppl.* **70**, 419–446 (1989).
24. Barnes, J. E. & Hut, P. *Nature* **324**, 446–449 (1986).
25. Hernquist, L. *Astrophys. J. Suppl.* **64**, 715–734 (1977).
26. Gingold, R. A. & Monaghan, J. J. *Mon. Not. R. astr. Soc.* **181**, 375–389 (1977).
27. Lucy, L. *Astr. J.* **82**, 1013–1024 (1977).
28. Barnes, J. E. & Hernquist, L. *Astrophys. J.* **370**, L65–68 (1991).
29. Barnes, J. E. *Astrophys. J.* **331**, 699–717 (1988).
30. Hernquist, L. *Nature* **340**, 687–691 (1989).
31. Malin, D. F. in *New Aspects of Galaxy Photometry* (ed. Nieto J.-L.) **27–37** (Springer, Berlin, 1985).
32. Heckman, T. M. *et al. Astrophys. J.* **311**, 526–547 (1986).
33. Vader, J. P., Heisler, C. A. & Frogel, J. A. *Ann. N. Y. Acad. Sci.* **571**, 247–248 (1989).
34. Toomre, A. in *The Evolution of Galaxies and Stellar Populations* (eds Tinsley, B. M. & Larson, R. B.) 401–426 (Yale University Observatory, New Haven, 1977).

ACKNOWLEDGEMENTS. This work was supported in part by an allocation of Cray time from the Pittsburgh Supercomputing Center, and by the Alfred P. Sloan Foundation, NASA, the California Space Institute and the NSF.

Merging of two galaxies with central black holes

T. Ebisuzaki, J. Makino* & S. K. Okumura

Department of Earth Science and Astronomy, and

* Department of Information Science and Graphics,

College of Arts and Sciences, University of Tokyo, 3–8–1 Komaba, Meguro-ku, Tokyo 153, Japan

OBSERVATIONS of elliptical galaxies show a positive correlation^{1,2} between total luminosity and core radius, defined as the radius at which surface luminosity becomes half of the central value. This has been taken as evidence against the idea that ellipticals are the result of galaxy mergers, because in simulated mergers^{3–5} the core radius remained almost constant. In those simulations, the galaxies contained no central black holes, but there is a body of evidence to suggest that such black holes are common in ellipticals⁶. Here, we present simulations of the merging of identical galaxies with and without central black holes, and find that when black holes are included, the merged galaxy acquires an isothermal core comparable in mass to the sum of the two initial black holes. Furthermore, the ratio of the core radius to the half-mass radius is approximately the same as the ratio of the black hole mass to the total galaxy mass, a result also consistent with observational evidence. These results, which can be understood by means of simple analytical arguments, suggest that most elliptical galaxies contain central black holes with masses comparable to the mass of their cores.

The hypothesis^{7,8} that most elliptical galaxies are formed by the merging of spiral galaxies is strongly supported by recent observations^{9–12}. *N*-body simulations have shown that merging of two spiral galaxies produces elliptical galaxies⁴. Such simulations have not, however, been able to reproduce the observed positive correlations between the core radius, r_c , and the total luminosity, L (refs 1, 2). The core radii remain almost constant during simulated merging of galaxies³. If giant elliptical galaxies

TABLE 1 Core radius

Model	N_0	Black hole	Before*	After†
1	512	yes	0.345	0.392
2	512	no	0.462	0.458
3	1,024	yes	0.376	0.474
4	1,024	no	0.451	0.421
5	2,048	yes	0.397	0.462
6	2,048	no	0.462	0.452

* Average from $t=5.00$ through 7.25

† Average from $t=37.50$ through 39.95.

were formed by multiple merging events, their cores would be almost the same size as those of fainter ellipticals. This constancy of the core radius is confirmed by analytical discussion based on conservation of central phase-space density¹³ and by recent numerical simulations⁵ with larger numbers of particles ($N=16,384$).

The difficulty is removed by taking into account a massive black hole at the centre of a galaxy. It has been suggested that many galaxies have black holes at their centres (examples are M87, M31, M32 and our Galaxy)⁶. Radio observations show that ~10% of elliptical galaxies have radio lobes or jets which are generally believed to be due to central black holes¹⁴. The suggested mass of such central black holes in elliptical galaxies is $10^{8-10} M_\odot$, which is comparable to their core mass. Begelman *et al.*¹⁵ and Rees¹⁶ investigated the evolution of a black-hole binary in a merger remnant. They did not, however, consider how gravitational energy release by the black holes changes the structure of the parent galaxies. To investigate the effect of massive black holes on the structure of the merger remnant, we performed *N*-body simulations of the merging of two identical galaxies with and without central black holes.

In our simulations with black holes, each initial galaxy is represented by N_0 particles with mass $m_i = 1/N_0$. Symbols with subscript 0 are the values before merging. We performed three

calculations with $N_0 = 512, 1,024$ and $2,048$ to investigate the effect of N_0 . We used the standard system of units¹⁷ in which $M_0 = 1$, $E_0 = -1/4$ and $G = 1$. The central black hole is represented by one particle of mass $M_{\text{BH},0} = 0.02$. We adopted the Plummer model with a half-mass radius $r_{h,0}$ of 0.77. The pericentric distance of the initial relative orbit of the galaxies was set to be 1.0. The simulations were performed on GRAPE-2, a special-purpose computer for gravitational N -body problems¹⁸. We used the NBODY3 code¹⁹ modified for GRAPE-2 (ref. 20).

Table 1 shows the core radii before and after merging. Here, we use the definition of r_c given by Casertano and Hut²¹. The 1σ errors of the core radii are typically $\sim 5\%$. For all three simulations with different numbers of particles, the core radii expand by ~ 14 – 25% if central black holes exist, whereas they remain almost constant if no black hole exists. No significant dependence of the core radius on $m_i/M_{\text{BH},0}$ can be seen. Figure 1a and b shows the density profiles (thick curve) of the merger remnants for the case of $N_0 = 1,024$ at $t = 40$. Thin curves in Fig. 1a and b show the density profile of the galaxy before merging. In Fig. 1a, the core radius, r_c , significantly increases and the core density decreases after merging, whereas in Fig. 1b, r_c remains almost constant and the density increases. Thus, the expansion of the core is caused by the gravitational energy release of a pair of black holes. The energy released by the

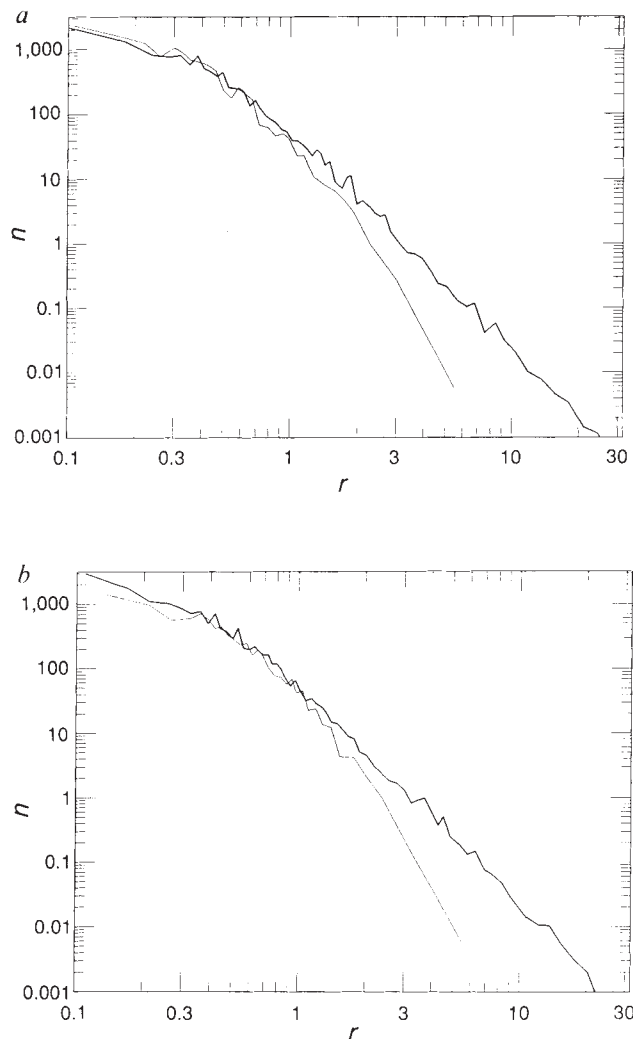


FIG. 1 Radial density profile of merger remnants (a) with and (b) without black holes. For comparison, the density profiles of the initial galaxies are shown by thin curves. The core expanded in the case of merging with black holes (a).

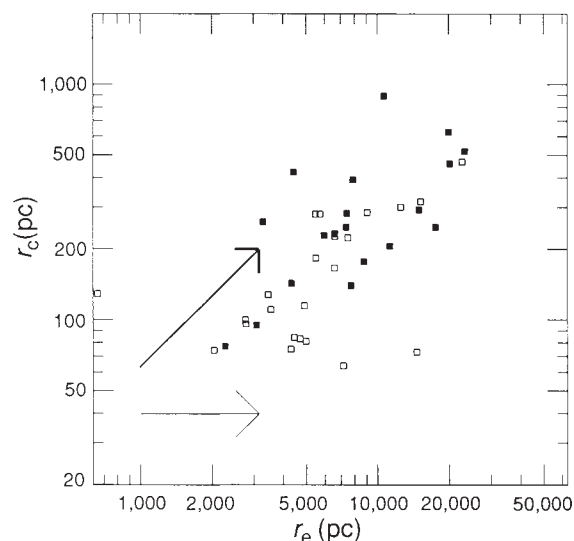


FIG. 2 Core radii plotted against the effective radii of elliptical galaxies obtained by Lauer². The filled squares denote well resolved galaxies (resolution class I and II) and open squares denote poorly resolved galaxies (resolution class III). Thick and thin arrows indicate the directions of evolution during merging with and without central black holes, respectively.

black-hole binary is $\sim 20\%$ of the total binding energy of the remnant. This is large enough to change the structure of the core. The results of simulations without black holes are in good agreement with previous calculations.

The evolution of the black holes is divided into three phases. First, black holes rapidly sink toward the common centre of the system with a timescale of $\tau_{\text{sink}} = (m_i/M_{\text{BH},0})\tau_r$ due to dynamical friction, where τ_r is the half-mass relaxation time of the galaxy¹⁵. The timescale of sinking, τ_{sink} , is as short as the crossing time for the case of $M_{\text{BH},0} = 0.02 M_0$. In this first phase, the mass, M_i , of the field particles between black holes is much higher than M_{BH} . Therefore, the distribution of the field particles is not affected by the black holes.

As the black holes sink, their mass becomes significant compared with M_i . After that, we may regard them as a self-bounded binary. The energy released by the black-hole binary is large enough to change the density distribution of field particles significantly. The field particles expelled by the black-hole binary are likely to produce a uniform density distribution. A black-hole binary in a uniform distribution of field particles loses angular momentum rather than energy, because dynamical friction is most effective at the apocentre, which has the lowest orbital velocity and the largest moment of inertia. In this second phase, the pericentric distance rapidly decreases although the apocentric distance remains almost constant.

Eventually, emission of the gravitational wave becomes significant at the pericentre, and two black holes merge. In this third phase, energy is released mainly in the form of the gravitational wave.

Begelman *et al.*¹⁵ investigated the evolution of a black-hole binary in a merger remnant. They assumed that the orbit of the binary remained circular. They concluded that orbital decay slows down when the separation of the binary decreases to ~ 0.1 – 0.01 of the core radius. But as discussed above, the orbit becomes highly eccentric. This is confirmed by numerical simulations of the orbital decay of a black-hole binary due to dynamical friction²². When gravitational waves become significant, the apocentric distance still remains ~ 0.1 – 0.01 of the core radius of the parent galaxies.

Black holes produce an isothermal core in the second phase of the evolution where $M_{\text{BH},0} \approx M_i$. Nonisothermal structures such as density cusps are destroyed. The mass, M_c , of the new

core is comparable to M_{BH} :

$$M_{\text{c}} \approx M_{\text{BH}} = 2M_{\text{BH},0} \quad (1)$$

Here M_{BH} is the mass of the black hole after merging. In our simulations, the core mass of the merger remnant is several times larger than M_{BH} . This is simply because the core mass of the initial galaxy is ~ 10 times as large as $M_{\text{BH},0}$. We adopted the model of an initial galaxy with a relatively large core to suppress the change in core radius due to two-body relaxation. The velocity dispersion in the core is almost equal to that around the half-mass radius, r_{h} . Using the virial theorem, we obtain

$$\frac{M_{\text{c}}}{r_{\text{c}}} \approx \frac{M}{r_{\text{h}}} \quad (2)$$

where M is the total mass of the remnant. In a dissipationless merging, the ratio of black-hole mass to total mass remains almost constant, as both the total mass and the black-hole mass almost double through the merging

$$\frac{M_{\text{BH}}}{M} \approx \frac{M_{\text{BH},0}}{M_0} \quad (3)$$

Using equations (1), (2) and (3), we obtain

$$\frac{r_{\text{c}}}{r_{\text{h}}} \approx \frac{M_{\text{BH}}}{M} \quad (4)$$

Therefore, $r_{\text{c}}/r_{\text{h}}$ should be almost constant among elliptical galaxies if most of these are formed by the merging of galaxies containing central black holes whose mass is several per cent of the total mass. This is consistent with the observations. In Fig. 2, we plot the core radius, r_{c} , against the effective radius, r_{e} of elliptical galaxies, using the data obtained by Lauer². Here, r_{h} is proportional to r_{e} . Thick and thin arrows in Fig. 2 indicate the direction of evolution during merging with and without black holes, respectively. The estimate described above is not very sensitive to the mass ratio of the black holes. We expect our result to be basically unchanged for mass ratios of ~ 2 –3.

We conclude that the core radius of the remnant is enlarged when galaxies with central black holes merge, because of the gravitational energy release of the black holes. The mass of the new core is comparable to that of the black hole, independent of its initial value. The ratio $r_{\text{c}}/r_{\text{h}}$ is close to M_{BH}/M , which is several per cent. The half-mass radius increases by a factor of ~ 1.4 –2.0 after a merging event;⁵ this has been confirmed by simulations with many more particles ($N = 16,384$) but without black holes. These results, together with the merger hypothesis, explain the observed correlation of core radius and half-mass radius with total luminosity^{1,2}. In contrast, the core mass of the remnants is smaller than that of the initial galaxies in the case of merging of galaxies without black holes. This suggests that most elliptical galaxies have central black holes. $\square$

19. Aarseth, S. J. *Multiple Time Scales* (eds Brackhill, J. U. & Cohen, L.) 377–418 (Academic, New York, 1985).
20. Makino, J. *Publ. Astr. Soc. Japan* (in the press).
21. Casertano, S. & Hut, P. *Astrophys. J.* **298**, 80–94 (1985).
22. Fukushima, T., Ebisuzaki, T. & Makino, J. *Publ. Astr. Soc. Japan* (submitted).

ACKNOWLEDGEMENTS. We thank D. Sugimoto for helpful discussions. We also thank T. Ito who built GRAPE-2 and S. J. Aarseth who kindly provided us with his code NBODY3. This work is partially supported by the Grant-in-Aid for General Scientific Research of The Ministry of Education, Science and Culture, Japan.

Near-field optical imaging with a non-evanescently excited high-brightness light source of sub-wavelength dimensions

Aaron Lewis & Klony Lieberman

Division of Applied Physics, The Hebrew University of Jerusalem, Jerusalem, Israel

NEAR-field optics involves scanning a spot of light, of dimensions smaller than a wavelength, across the surface of a sample at a distance small enough (a few hundred ångströms) that far-field diffraction effects do not occur^{1,2}. In principle this method should generate an image with a resolution determined principally by the dimensions of the spot of light and not limited by the wavelength³. Although near-field imaging was first proposed in 1928^{4,5}, efficient implementations that overcome the problem of large evanescent losses in passing light through a sub-wavelength aperture have not been previously realized. Here we present a technique that creates a point of sub-wavelength light without associated evanescent losses in its excitation while achieving the advantage of the exponential increase in intensity that occurs within the near-field³. The sub-wavelength light source is provided by a micropipette coated with a metal and filled with a fluorescent dye embedded in a plastic matrix. The instrument we describe combines the potential for near-field microscopy with the characteristics of a conventional far-field light microscope. Images with overlapping resolutions can be obtained having magnifications that range from a few hundred with the conventional microscope to magnifications, in the near-field mode, of tens of thousands that are of the order normally associated with scanning electron microscopy. Such imaging with light can be achieved even with fluorescence, under ambient conditions and without the destructive sample preparation and beam damage that is characteristic of electron microscopy.

The method of generating a sub-wavelength spot of light is an important aspect of near-field imaging. The first approach used was to impress a small sub-wavelength aperture in a flat metal screen and to use it as an aperture for a larger light source⁶. Although the method was used for imaging high-contrast objects⁷, images could only be generated of perfectly flat surfaces because of the difficulty of positioning such an aperture close enough to all regions of a rough sample. An approach that could overcome this problem was to place the aperture at the tip of a sharp rod. Pohl and co-workers⁸ etched a single crystal of quartz to a sub-wavelength tip, which was coated with metal and deformed at the tip to produce the aperture. An alternate approach⁸ exploited the fact that glass pipettes used routinely for electrophysiology can be produced with apertures at the tip as small as a few hundred ångströms. These pipettes can be generated using a commercial puller based on a filament as the heating element. By coating such a pipette with metal, a sub-wavelength aperture can be created. A related technique, developed by several workers^{6,10–12}, detects alterations in an evanescent field generated on a surface and extending to only a few hundred ångströms above the surface. The method has been successfully implemented^{10–12} by using an etched fibre

Received 27 February; accepted 8 October 1991.

1. Kormendy, J. *Astrophys. J.* **295**, 73–79 (1985).
2. Lauer, T. R. *Astrophys. J.* **292**, 104–121 (1985).
3. Farouki, R. T., Shapiro, S. L. & Duncan, M. J. *Astrophys. J.* **265**, 597–605 (1983).
4. Barnes, J. E. *Astrophys. J.* **331**, 699–717 (1988).
5. Okumura, S. K., Ebisuzaki, T. & Makino, J. *Publ. Astr. Soc. Japan* **43** (in the press).
6. Dressler, A. *IAU Symp. No. 134, Active Galactic Nuclei* (eds Osterbrock, D. E. & Miller, J. S.) 217–232 (1989).
7. Toomre, A. & Toomre, J. *Astrophys. J.* **178**, 623–666 (1972).
8. Toomre, A. *Evolution of Galaxies and Stellar Populations* (eds Tinsley, B. M. & Larson, R. B.) 401–426 (Yale Observatory, New York, 1977).
9. Schweizer, F. *Astrophys. J.* **252**, 455–460 (1982).
10. Bergvall, N., Rönnebeck, J. & Joansson, L. *Astr. Astrophys.* **222**, 49–57 (1989).
11. Wright, G. S., James, P. A., Joseph, R. D. & McLean, I. S. *Nature* **344**, 417–419 (1990).
12. Kormendy, J. & Djorgovski, S. *Ann. Rev. Astr. Astrophys.* **27**, 235–277 (1989).
13. Carlberg, R. L. *Astrophys. J.* **310**, 593–596 (1986).
14. Sadler, E. M. *Astr. J.* **89**, 53–63 (1984).
15. Begelman, M. C., Blandford, R. D. & Rees, M. J. *Nature* **287**, 307–309 (1980).
16. Rees, M. J. *Science* **247**, 817–823 (1990).
17. Heggie, D. C. & Mathieu, R. D. in *The Use of Supercomputers in Stellar Dynamics* (eds McMillan, S. & Hut, P.) 233–235 (Springer, Berlin 1986).
18. Ito, T., Makino, J., Ebisuzaki, T. & Sugimoto, D. *Publ. Astr. Soc. Japan* **43**, 547–555 (1991).

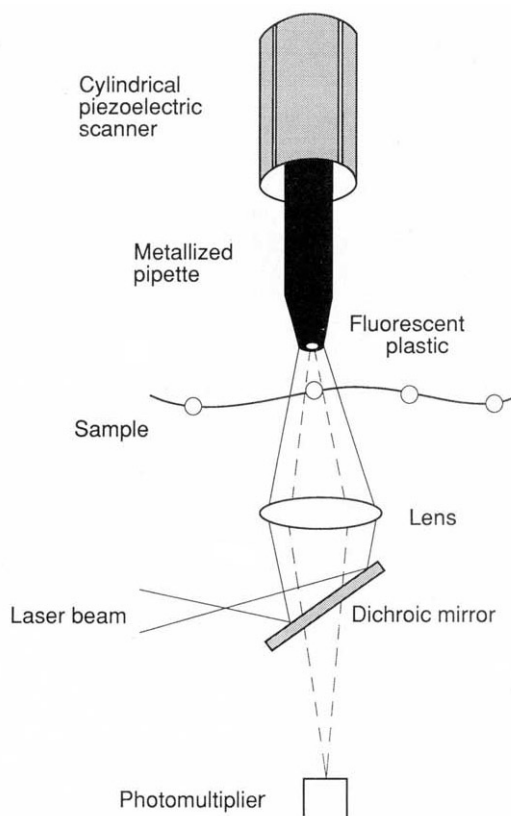


FIG. 1 Schematic diagram of the method used to produce intense sub-wavelength spots of light non-evanescently for high-resolution imaging. A laser beam (solid line) is reflected off a dichroic mirror and focused by a lens onto the tip of a glass micropipette which is coated with metal on the outside. The pipette tip is filled with a fluorescent plastic, and the light that is emitted from this plastic tip (dotted lines) is transmitted through the surface to be imaged and through the dichroic mirror onto a photomultiplier.

which produces sub-wavelength topographical images of surface features. Recently, Betzig *et al.*¹³ have introduced the use of a single-mode optical fibre pulled to a tip, by heating with a CO₂ laser instead of a filament. The fibre is coated with aluminium to produce apertures as small as 20 nm.

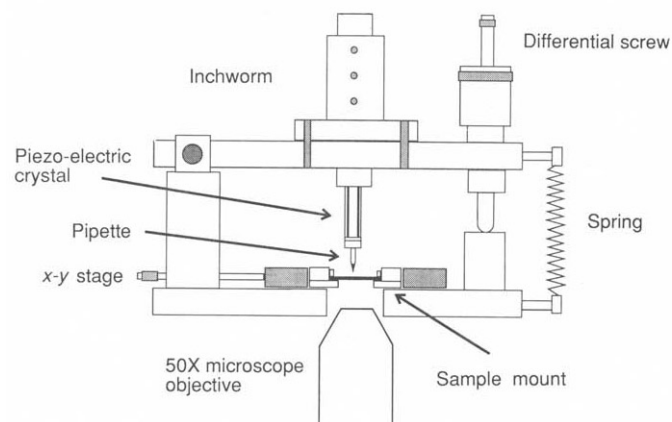
All of the above methods suffered from the problem that in a sub-wavelength aperture, there are no propagating modes. The light wave enters a region of evanescence and its intensity decays exponentially. A way around this problem was demonstrated by Lieberman *et al.*¹⁴ who were able to grow a molecular crystal in the tip of the pipette. The light transmitted through the metal-coated pipette was converted by the crystal into molecular excitons which could traverse even sub-wavelength regions and

would emit light after travelling an appropriate distance. Experimental results indicated that the efficiency of light transmission through such apertures increased threefold. The relatively small improvement in the throughput arose for two reasons. First, the molecular crystals used were produced from anthracene which underwent considerable photochemical oxidation. Second, the maximum pathlength for exciton diffusion in the crystal is thought to be a few micrometres¹⁵, whereas the region of evanescence in pipettes with the best profiles could be significantly longer than the exciton diffusion length. Thus, in spite of this advance, the light intensities obtained were not sufficient for low-contrast imaging.

Our approach is schematically described in Fig. 1. In this method, a metal-coated pipette tip is filled with a highly stable form of the fluorescent dye pyrene (BASF 241, with a maximum wavelength of absorption at 525 nm and an emission at 540 nm) embedded in a plastic matrix. The tip is formed by dissolving the dye in a 2% solution of polymethylmethacrylate in chloroform, and injecting the solution into the pipette where it is allowed to solidify. The tip of the pipette is then excited head-on with a laser beam reflected off a standard dichroic mirror, producing an intense sub-wavelength light source. Because emission occurs at the very tip of the pipette, the evanescent losses in the excitation source are minimized. The best results yet obtained from apertured sources have been 50 nW for an 80-nm aperture⁸; in contrast our method of excitation can produce intensities of 5 μ W for a 100-nm aperture, using a 514.5-nm argon ion laser with 8×10^{-4} W μ m⁻² incident on the plastic tip. With such intensities we have seen no effects of bleaching in the light source. This results from both the stability of the dye and the isolation of the molecules by the surrounding plastic matrix. Such a method of excitation is readily integrated into a conventional fluorescence microscope, and permits considerable flexibility in combining methods that allow feedback in the aperture/sample separation which must be maintained to obtain the highest near-field resolution. Another advantage of the method is that the excitation of the pipette tip is mechanically uncoupled from the near-field microscope. In previous implementations of near-field microscopy the light is either brought to the surface or collected from the surface by a fibre which may couple mechanical vibrations into the microscope. The diagram in Fig. 1 assumes a standard inverted microscope equipped with a dichroic mirror which reflects a laser beam through the objective and the sample onto the pipette tip. The fluorescent light emitted passes through the sample and is collected by the objective.

To scan the pipette light source over the surface, we replaced the sample stage of the Zeiss Axiovert microscope used in our measurements with a design conventionally used in scanning tunnelling microscopy¹⁶ (Fig. 2). The pipette is held in a conventional tube scanner used in scanning tunnelling microscopy for three-dimensional fine scanning. The pipette/tube scanner is inserted into a Burleigh inchworm for a coarse approach to the surface. The tip assembly opens up for rapid replacement of fluorescing tips.

FIG. 2 Schematic diagram of the near-field microscope, which is positioned in place of the sample stage of the conventional far-field inverted microscope represented by a microscope objective in this diagram. The pipette is held in a conventional tube scanner used in scanning tunnelling microscopy for three-dimensional fine scanning. The pipette/tube scanner is inserted into a Burleigh inchworm for a coarse approach to the surface. The tip assembly opens up for rapid replacement of fluorescing tips.



The sample and the pipette can be viewed during this approach with the inverted microscope, allowing the pipette tip to be quickly located over the region of the sample for which a high-resolution near-field image is to be obtained. The light collected by the objective of the microscope is focused through the dichroic mirror onto an uncooled Hitachi photomultiplier connected to a personal computer for image processing.

Figure 3 shows an image obtained of low-contrast, unstained latex spheres $0.9\ \mu\text{m}$ in diameter, using a plastic-tipped pipette illuminated with 100 mW of the 488.0-nm line of an argon ion laser focused to a $50\ \mu\text{m}^2$ spot with a long-working-distance objective ($\times 50$, numerical aperture 0.4) onto the front face of a $0.5\text{-}\mu\text{m}$ diameter pipette. This image consists of 128×128 pixels, and the signal is integrated for 20 ms at each point. The pipette is approaching the near-field but is still at a distance of over one wavelength from the sample, as is evident from the well-defined Fresnel rings seen around each of the spheres. Because of their large radius of curvature, points on the balls illuminated by the laser that were out-of-focus relative to the near-field of the pipette tip apparently resulted in constructive interference.

Figure 4 shows an image of a suspension of $0.04\text{-}\mu\text{m}$ uniform gold balls dispersed on a glass slide. The same illumination conditions were used, but this image (256×256 pixels) was obtained with a $0.25\text{-}\mu\text{m}$ diameter pipette and a 1-ms integration time at each pixel. The reduction in the integration time was made possible by the highly scattering nature of the balls. As the sample was essentially flat we were able to approach into the near field, and thus no Fresnel rings are seen. Groups of balls resolved in this image appear as single scattering objects in a far-field image (not shown). The image resolution in this case is determined by the pipette diameter. Such balls can be attached to a variety of biological molecules¹⁷ and could be used to enhance contrast in this approach to near-field scanning optical microscopy as is the case in electron microscopy. Further theoretical work is needed to interpret fully the effects of absorption, scattering and quenching of the fluorescent tip.

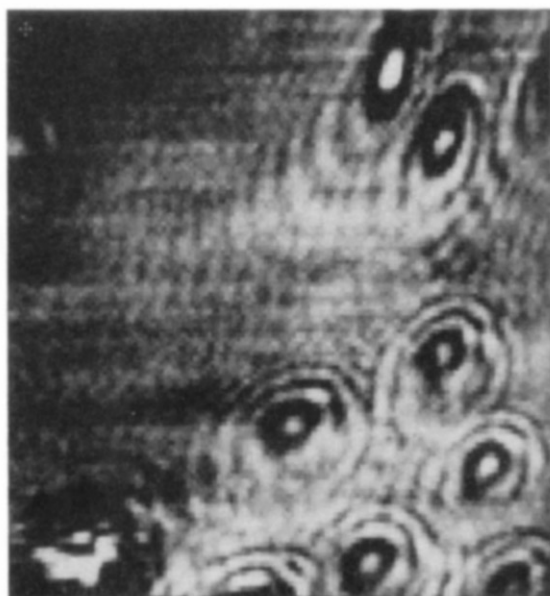


FIG. 3 An image of a region of the field of $0.9\text{-}\mu\text{m}$ latex spheres taken with the fluorescent tip. The region displayed in this image is $12 \times 12\ \mu\text{m}$. The tip is more than one wavelength from the sample, as we observe well-defined Fresnel diffraction.

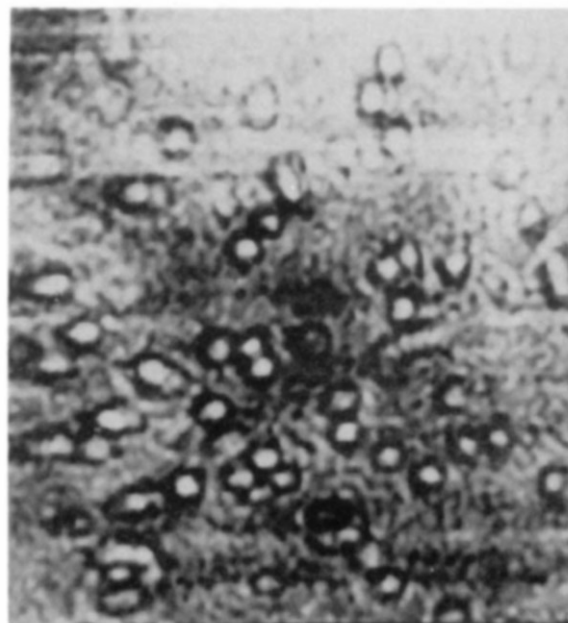


FIG. 4 A near-field image of a $12 \times 12\ \mu\text{m}$ field of $0.04\text{-}\mu\text{m}$ gold balls distributed on a glass slide. The resolution obtained was determined by the $0.25\text{-}\mu\text{m}$ pipette diameter. These balls have been widely used as markers of biological molecules.

The intensity of the fluorescent point source used here scales linearly with the area rather than decreasing exponentially when light has to be transmitted or collected through a sub-wavelength pipette or fibre. Thus there is no reason why the size of this spot of light cannot be reduced to a few nanometres (although below this size the metal coating of the pipette will quench the fluorescence and reduce the intensity of the source¹⁸). Because of the method of excitation of the sub-wavelength light spot and the fact that this excitation is mechanically uncoupled from the near-field microscope, the microscope construction is simplified, and it should be possible to combine this approach with feedback mechanisms such as force sensing². Finally, this excitation method enhances the contrast of the sample, because the pipette acts both as a detector of variations in the excitation and as a source of imaging light. □

Received 24 April; accepted 7 August 1991.

1. Pohl, D. W. *Advances in Optical and Electron Microscopy* Vol. 12, 243–312 (Academic, London, 1991).
2. Lewis, A. & Lieberman, K. *Analyt. Chem.* **63**, A625–638 (1991).
3. Betzig, E., Harootunian, A., Lewis, A. & Isaacson, M. *Appl. Opt.* **25**, 1890–1900 (1986).
4. Synge, E. H. *Phil. Mag.* **6**, 356–362 (1928).
5. McMullan, D. *Proc. R. microsc. Soc.* **25**, 127–131 (1990).
6. Lewis, A., Isaacson, M., Harootunian, A. & Murray, A., *Biophys. J.* **41**, A405 (1983); *Ultramicroscopy* **13**, 227–232 (1984).
7. Fischer, U. C. *J. Vac. Sci. Technol.* **B3**, 386–390 (1985).
8. Pohl, D. W., Denk, W. & Lanz, M. *Appl. Phys. Lett.* **44**, 651–653 (1984).
9. Harootunian, A., Betzig, E., Isaacson, M. & Lewis, A. *Appl. Phys. Lett.* **49**, 674–676 (1986).
10. Reddick, R. C., Warmack, R. J. & Ferrell, T. L. *Phys. Rev.* **B39**, 767–770 (1989).
11. Courjan, D., Sarayeddine, K. & Spayer, M. *Opt. Commun.* **71**, 23–28 (1989).
12. DeFornel, F., Goudonnet, J. P., Salomon, L. & Leshiewska, E. *SPIE* **1139**, 77–84 (1989).
13. Betzig, E., Trautman, J. K., Harris, T. D., Weiner, J. S. & Kostelak, R. L. *Science* **251**, 1468–1470 (1991).
14. Lieberman, K., Haroush, S., Lewis, A. & Kopelman, R. *Science* **247**, 59–61 (1990).
15. Kopelman, R. *Science* **241**, 1620–1626 (1988).
16. Hansma, P. K., Elings, V. B., Marti, J. & Bracker, C. E. *Science* **242**, 209–216 (1988).
17. Geertz, H., De Brabander, M., Nuydens, R. & Nuyens, R. in *New Techniques in Optical Microscopy and Microspectroscopy* (ed. Cherry, C.) (Macmillan, London, 1991).
18. Drexhage, K. *Prog. Opt.* **12**, 163–232 (1974).

ACKNOWLEDGEMENTS. We thank R. Grishi and R. Reisfeld for the BASF dye. The Sutter Instrument Company is gratefully acknowledged for a P2000 CO₂ laser puller for pulling quartz capillaries. This work was supported by the United States Air Force, by the United States-Israel Binational Science Foundation and by the Israel National Council for Research and Development.

Stabilization of superionic α -AgI at room temperature in a glass matrix

Masahiro Tatsumisago, Yoshikane Shinkuma
& Tsutomu Minami

Department of Applied Chemistry, University of Osaka Prefecture, Sakai,
Osaka 591, Japan

SINCE the discovery¹ that the high-temperature phase of silver iodide (α -AgI) has an ionic conductivity comparable to that of the best liquid electrolytes, solid electrolytes have attracted wide interest. Possible applications of these materials range from solid-state batteries to electrochromic displays and sensors². Although α -AgI displays conductivities of more than 10 S cm^{-1} (ref. 3), owing to the almost liquid-like mobility of Ag^+ ions, the crystal transforms below 147°C to the β -phase with a conductivity of only $\sim 10^{-5}\text{ S cm}^{-1}$ at room temperature. Efforts to achieve good conductivities at lower temperatures have focused on the addition of a second component to AgI to form solid solutions or new compounds such as RbAg_4I_5 and Ag_2HgI_4 (refs 4–7). Here we report our success in depressing the $\alpha \rightarrow \beta$ transformation temperature so as to stabilize α -AgI itself at room temperature. We use a melt-quenching technique to prepare crystallites of α -AgI frozen into a silver borate glass matrix. The quenched material showed diffraction peaks characteristic of α -AgI and displayed ionic conductivities of about 10^{-1} S cm^{-1} . Further development of these glass/crystal composites may make the high ionic conductivity of α -AgI available for room-temperature solid-state applications.

We have developed AgI-based glassy materials, which have high ionic conductivities (10^{-3} – 10^{-2} S cm^{-1}) at ambient temperature, by combining AgI and silver oxosalts or oxides^{8–10}.

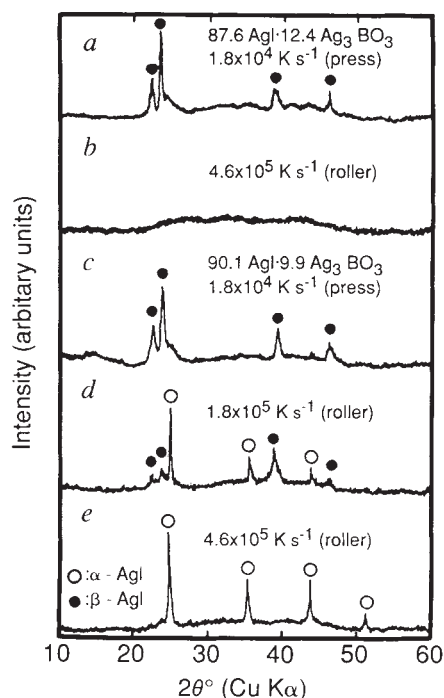


FIG. 1 X-ray diffraction patterns for the melt-quenched samples with different cooling rates and compositions in the system $\text{AgI-Ag}_3\text{BO}_3$. 'Press' and 'roller' indicate metal-plate press quenching and twin-roller quenching, respectively.

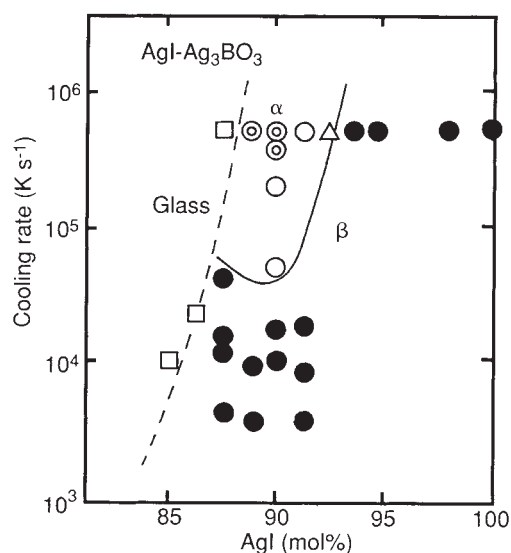


FIG. 2 Crystalline phases observed in the melt-quenching samples with various cooling rates and compositions in the system $\text{AgI-Ag}_3\text{BO}_3$. $\square$: glass, $\circ$: α -AgI only, $\triangle$: β -AgI > α -AgI, $\bullet$: β -AgI only.

We refer to such materials as 'superionic conducting glasses'¹¹. When a very high proportion of AgI was used in the combination of starting materials to be melted, a small amount of crystalline AgI was found to be precipitated in the glassy phases. The precipitated AgI was generally reported not to be the α -phase but the less-conductive β -phase¹².

In contrast to this, we found that the α -AgI phase was frozen in rapidly quenched glassy matrices in the course of preparing superionic conducting glasses containing large amounts of AgI to achieve extremely high conductivities. Here we describe in detail the experimental conditions needed to obtain the high-temperature α -AgI phase in glass matrices. The pseudobinary system $\text{AgI-Ag}_3\text{BO}_3$ was chosen as the glass-forming system, and we also show the characterization of such novel composite materials, composed of α -AgI and $\text{AgI-Ag}_3\text{BO}_3$ glass matrices.

Reagent-grade AgI, Ag_2O and B_2O_3 were used as starting materials for sample preparation¹³. Batches (2 g) of the mixed materials were melted at 500 – 600°C for 30 min in a silica glass tube (10 mm in diameter) with one open end. We used two melt-quenching techniques: metal-plate press quenching and twin-roller quenching^{14,15}. In the former, the melt was poured onto a stainless steel plate and pressed with another plate into a thin disk (~ 0.1 – 0.5 mm thick). In the latter, the melt was poured into the rotating twin rollers (3,000 r.p.m.) to yield thin flakes (~ 0.01 – 0.05 mm thick). We estimated the quenching rate from the thickness of the products using the equations of Sarjeant¹⁶ and Nakamura¹⁷.

X-ray diffraction measurements were carried out with the flakes or disk glued on a glass plate. The conductivity was measured for the flakes, on which two parallel gold electrodes were evaporated ~ 0.5 mm apart. The measurement was carried out in dry nitrogen using an impedance analyser (HP 4192A) in the temperature range 25 – 250°C (ref. 18).

Figure 1 shows the X-ray diffraction patterns at ambient temperature for samples in the system $\text{AgI-Ag}_3\text{BO}_3$ with different AgI contents and different cooling rates. As glass formation was observed in the range 57 – 81 mol% AgI when the metal-plate press quenching was performed, the composition in Fig. 1 is out of the glass-forming region. In the case of 87.6 mol% AgI, the diffraction peaks due to β -AgI are only observed for metal-plate press quenching (Fig. 1a). For twin-roller quenching, a halo pattern due to the glassy state is observed (Fig. 1b)

for this composition. In the case of 90.1 mol% AgI, metal-plate press quenching produces only β -AgI (Fig. 1c), as in the case of 87.6 mol% AgI. When twin-roller quenching is used, however, the diffraction peaks due to α -AgI are clearly observed (Fig. 1d), and the peak intensity increases with increasing cooling rate, as shown in Fig. 1e. It is thus obvious that both the composition and the melt-cooling rate are very important for obtaining α -AgI at ambient temperature.

Figure 2 is a phase diagram to show which phases are obtained when the cooling rate and the chemical composition are changed. Different symbols represent different phases obtained during melt quenching. Open squares denote the glassy phase. Double circles mean that only the α -AgI phase was observed. Open circles and open triangles indicate the co-existence of α - and β -phases; the α -phase is dominant in open circles and the β -phase in open triangles. Closed circles denote that only the β -AgI phase was observed. It is seen that the α -AgI phase is frozen only when the AgI content is slightly larger than that in the glass-forming limit. The cooling rate should also be higher than 10^5 K s^{-1} to obtain α -AgI at ambient temperature. Although the $\alpha \rightarrow \beta$ transformation in AgI crystal morphology occurs extremely rapidly¹⁹, we suggest that the increase in the viscosity of the matrix melts (undercooled liquids) during cooling prevents the $\alpha \rightarrow \beta$ transformation of the precipitated α -AgI to the β -phase; the cooling rate, and thus the rate of increasing viscosity, is very large in the case of twin-roller quenching.

Figure 3 shows the temperature dependence of conductivity for a twin-roller-quenched sample of 90.1 AgI·9.9 Ag₃BO₃ in which α -AgI is frozen in the matrices. The conductivity of an AgI crystal flake prepared by twin-roller quenching of pure AgI crystal is also shown for comparison. The conductivity at ambient temperature for samples with frozen-in α -AgI is $\sim 10^{-1} \text{ S cm}^{-1}$; such an extremely high ion conductivity is expected if composites composed of frozen α -AgI and superionic conducting glasses are formed. The activation energy in the temperature range 25 to 140 °C was 15 kJ mol⁻¹. These values are reasonable if the frozen α -AgI phase is present as highly dispersed, discrete particles, the conductivity of which was

calculated by Maxwell's equation for the conductivity of composite materials²⁰. With increasing temperature, a small conductivity change is observed near the $\alpha \rightarrow \beta$ transformation temperature of 147 °C, which is probably due to small amounts of residual β -phase. With decreasing temperature, however, no change in conductivity is observed at 147 °C, and the conductivity starts to decrease considerably at ~ 110 °C. The conductivity decrease is due to the $\alpha \rightarrow \beta$ transformation of the AgI crystal in the glass matrices; the cooling runs of high-temperature X-ray diffraction measurements for this sample confirm that this transformation occurs. This result indicates that even if the cooling rate is small, as in the conductivity measurements, glass matrices with high viscosity depress the $\alpha \rightarrow \beta$ transformation. Preliminary experiments for glass matrices other than borates indicate that a viscosity effect is likely; the details will be reported elsewhere. □

Received 13 June; accepted 26 September 1991.

1. Tubandt, C. & Lorenz, E. Z. *phys. Chem.* **24**, 513–543 (1914).
2. van Gool, W. (ed.) *Fast Ion Transport in Solids* (North-Holland, Amsterdam, 1973).
3. Funke, K., *Prog. Solid State Chem.* **11**, 345–402 (1976).
4. Bradley, J. N. & Greene, P. D. *Trans. Faraday Soc.* **63**, 424–430 (1967).
5. Suchow, L. & Pond, G. R. *J. Am. chem. Soc.* **75**, 5242–5244 (1953).
6. Takahashi, T., Ikeda, S. & Yamamoto, O. *J. electrochem. Soc.* **120**, 647–651 (1973).
7. Takahashi, T., Ikeda, S. & Yamamoto, O. *J. electrochem. Soc.* **119**, 477–482 (1972).
8. Minami, T., Nambu, Y. & Tanaka, M. *J. Am. Ceram. Soc.* **60**, 467–469 (1977).
9. Minami, T., Imazawa, Y. & Tanaka, M. *J. non-cryst. Solids* **42**, 469–476 (1980).
10. Minami, T. *J. non-cryst. Solids* **95/96**, 107–118 (1980).
11. Minami, T., Takuma, Y. & Tanaka, M., *J. electrochem. Soc.* **124**, 1659–1662 (1977).
12. Wnetrzewski, B., Nowinski, J. L. & Jakubowski, W. *Solid State Ionics* **36**, 209–211 (1989).
13. Minami, T., Ikeda, Y. & Tanaka, M. *Nippon Kagaku Kaishi*, 1617–1623 (1981).
14. Tatsumisago, M. & Minami, T. *Mater. Chem. Phys.* **18**, 1–17 (1987).
15. Sarjeant, P. T. & Roy, R. *J. Am. Ceram. Soc.* **50**, 500–503 (1967).
16. Sarjeant, P. T. & Roy, R. *Mater. Res. Bull.* **3**, 265–280 (1968).
17. Nakamura, T. & Takashige, M. *Kotaibutsuri* **15**, 497–503 (1980).
18. Tatsumisago, M., Machida, N. & Minami, T. *J. Ceram. Soc. Japan* **95**, 59–63 (1987).
19. Burley, G. *Acta Cryst.* **23**, 1–5 (1967).
20. Kingery, W. D., Bowen, H. K. & Uhlmann, D. R. *Introduction to Ceramics* 634–643 (Wiley, New York, 1975).

ACKNOWLEDGEMENTS. We thank T. Saito and K. Motomura for technical assistance, and acknowledge support for this research by a Grant-in-Aid for Scientific Research on Priority Areas, New Functional Materials—Design, Preparation and Control, the Ministry of Education, Culture and Science of Japan.

Effect of defects on molecular mobility in liquid water

Francesco Sciortino*, Alfons Geiger†
& H. Eugene Stanley*

* Center for Polymer Studies and Department of Physics, Boston University, Boston, Massachusetts 02215, USA

† Physikalisches Chemie, Fachbereich Chemie der Universität Dortmund, Postfach 500500, D-4600 Dortmund 50, Germany

LIQUID water is a totally connected random network of hydrogen bonds, the connectivity lying well above the percolation threshold^{1–3}. But despite this extensive association of hydrogen bonds with strengths greater than the thermal energy, the diffusion and rotation rates of water molecules at ambient temperatures are comparable to those of non-associated simple liquids. Many experiments have indicated that the random tetrahedral network cannot be perfect but must contain defects, which are characterized geometrically by the presence of an extra (fifth) molecule in the first coordination shell, or topologically by the presence of 'bifurcated' hydrogen bonds^{4–7}. Here we use molecular-dynamics simulations to examine the effect of such defects on molecular mobility in water. We find that they provide pathways of lower energy between different tetrahedral local arrangements, thus acting as 'catalysts'. The anomalous mobility of water under compression^{8,9} and the decreased mobility in hydrophobic hydration shells^{10,11} can be interpreted on the same basis. We suggest that our results are relevant to studies on 'stretched' water^{12,13}.

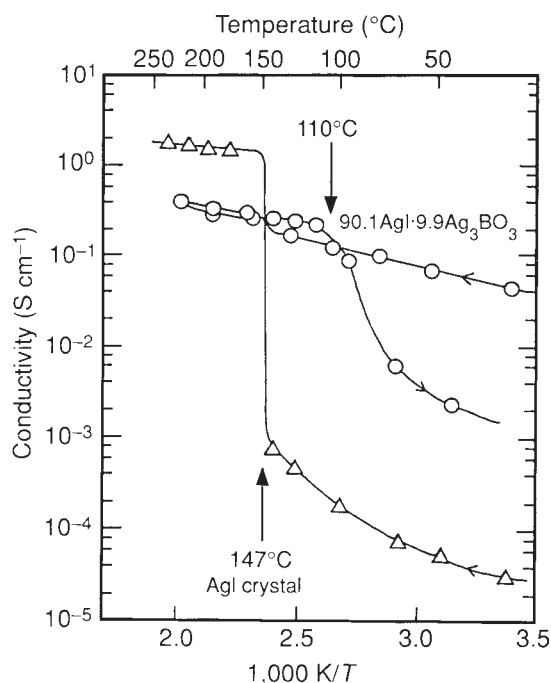


FIG. 3 Temperature dependence of conductivity for the twin-roller quenched samples of 90.1 AgI·9.9 Ag₃BO₃, in which α -AgI is the main component of the crystals, and for a crystal of AgI.

We have used two parallel approaches. The first is to change the global concentration of defects in the system: we tune the defect concentration not by changing the temperature but by changing the density of the simulated system¹⁴. This method enables us to distinguish between contributions to the mobility arising from thermal effects and those related to the defect concentration. The second approach is to compare the properties of groups of molecules that all belong to the same system but have different numbers of nearest neighbours. Using these two approaches, we can relate global behaviour to a microscopic picture based on the 'catalytic activity' of extra molecules.

In both approaches, we analyse quenched configurations, which represent the inherent structure of the liquid¹⁵. Because only local minima survive the quench process, the quenched

state represents the structure of the liquid without vibrational distortions. By analysing the bonding and structure after these thermal vibrations have been subtracted, we are able to make a clear distinction between purely thermal displacements and 'inherent' defects¹⁶, and to separate the effects on molecular mobility owing to network topology from those resulting simply from thermal excitation.

We study changes in local structure as a function of the global density of the system by looking at the changes in the first coordination shell. In Fig. 1a we show the average number of molecules $\langle n(r_{\min}) \rangle$ within a sphere of radius $r_{\min} = 3.3 \text{ \AA}$ as a function of the average density. Here $r_{\min}$ is the position of the first minimum in the radial distribution function. At density $\rho = 1 \text{ g cm}^{-3}$, there are ~ 4.6 neighbours; as ρ is lowered there is a progressive decrease until a value of 4.0 is reached at $\rho \approx 0.85$.

The average binding energy of the water molecules also decreases (Fig. 1b) as ρ is decreased. Thus Fig. 1a indicates that progressive 'stretching' of water leads to more tetrahedral bonding, and Fig. 1b demonstrates that stretching strengthens bonding.

Even at microscopic scales, a higher (less negative) binding energy is associated predominantly with a higher local density. To see this, we first calculate $P(E_i)$, the histogram of number of molecules with total binding energy E_i (Fig. 2a). For each energy bin, we then calculate the average coordination number $\langle n(r_{\min}) \rangle_{E_i}$ (Fig. 2b). We find that the average coordination number is enhanced for energies above the 'tetrahedral network' peak in $P(E_i)$, at $\sim -115 \text{ kJ mol}^{-1}$, demonstrating that high-energy states (above -115 kJ mol^{-1}) are associated with a denser local environment.

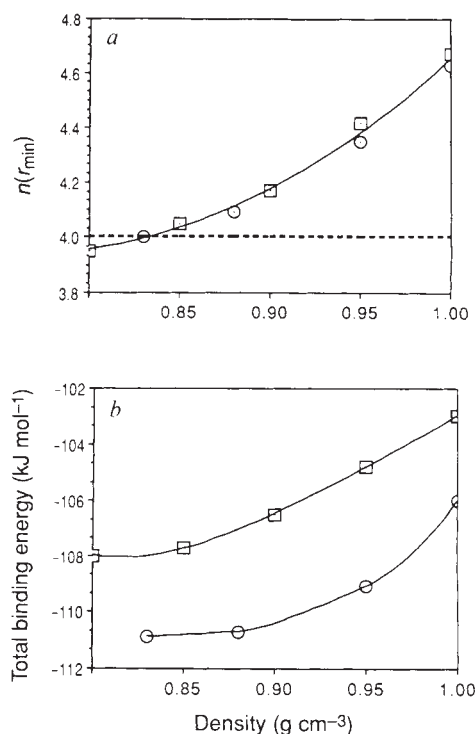


FIG. 1 Density dependence in ST2 water at 273 K ($\square$) and 235 K ($\circ$) for *a*, the average number of neighbours inside a shell of radius $r_{\min} = 3.3 \text{ \AA}$, and *b*, the average total binding energy $E_i = \sum_j V_{ij}$; here V_{ij} is the pair interaction energy between molecules i and j . Both quantities are calculated using quenched configurations. Our simulations concern a system of 216 water molecules interacting via the ST2 potential²³ (we found similar results for SPC²⁴ and TIP4P²⁵ potentials). Thus we perform simulations at a constant density and constant energy (the NVE ensemble). The direct molecular interactions up to the cutoff distance $r_c = 7.8 \text{ \AA}$ are combined with a reaction field approximation for more distant pairs. The density is decreased in steps from $\rho = 1.0 \text{ g cm}^{-3}$ to 0.8 g cm^{-3} . Two parallel simulation series were performed, with average temperatures of 273 and 235 K. Before the actual simulations were started, equilibration runs (extending up to 300 ps at low temperatures or densities) were applied. For a further description of the computational techniques, see ref. 21. We choose to study the 'low' temperature of $T = 235 \text{ K}$ to facilitate the recognition of effects related to the temperature. From each simulation we extract 50–100 equally spaced configurations (corresponding to a time interval between configurations ranging from 0.2 to 0.6 ps). We then calculate the respective 'quenched' (or 'inherent') configurations using a two-stage procedure¹⁵, (1) the equations of motion are strongly damped to remove the kinetic energy quickly; (2) the molecules are moved along the potential-energy surface toward the local minimum by means of a steepest-descent numerical calculation. This steepest-descent procedure is performed until the change in potential energy between successive steps is less than $10^{-5} \text{ kJ mol}^{-1}$ (we tested our results by performing a calculation using $10^{-6} \text{ kJ mol}^{-1}$, and we found essentially the same results).

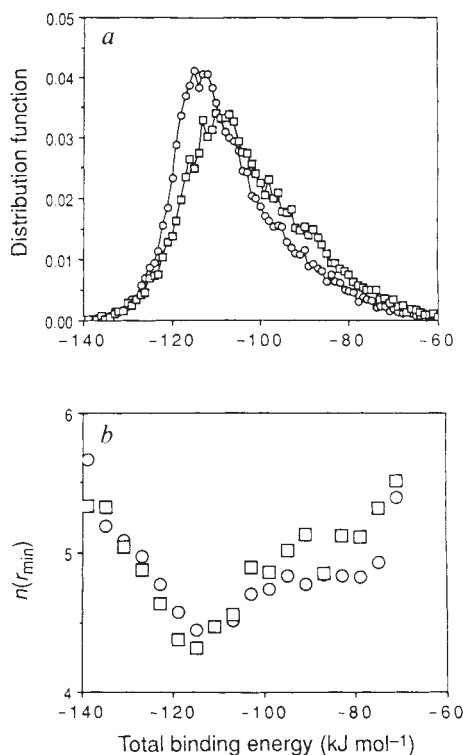


FIG. 2 *a*, Distribution function $P(E_i)$ for the binding energy $E_i = \sum_j V_{ij}$ calculated using the quenched configurations. Symbols: $T = 273 \text{ K}$ ($\square$) and 235 K ($\circ$). The average density here is $\rho = 1.0 \text{ g cm}^{-3}$. *b*, $\langle n(r_{\min}) \rangle_{E_i}$, the number of neighbours inside a shell of radius $3.3 \text{ \AA}$, averaged over all the molecules with the same E_i , as a function of the binding energy E_i . Note the enhancement for energies below, as well as above, the central 'perfect tetrahedral network' peak in $P(E_i)$ at $\sim -115 \text{ kJ mol}^{-1}$. The total potential energy of the system is, with our definition, $\frac{1}{2} \sum_i E_i$, as each interaction V_{ij} is counted twice.

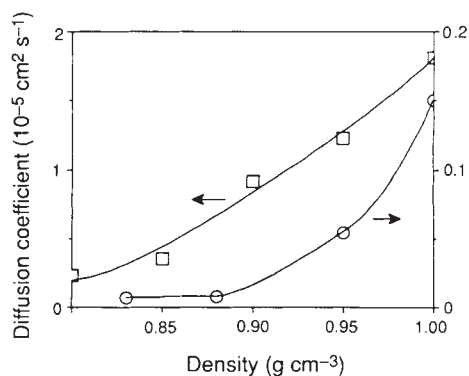


FIG. 3 Density dependence in ST2 water at 273 K ($\square$) and 235 K ($\circ$) for the diffusion coefficient. The diffusion coefficients are calculated from the long-time limit of the slope of a graph of mean square displacement against time.

An additional important observation from Fig. 2 is that there are a few very sparsely populated low-energy states (below -115 kJ mol^{-1}) that possess a high local density. An extra molecule can therefore fit in the first coordination shell in such a way as to decrease (that is, strengthen) the total bonding energy. This can be understood if there exists a 'bonded state' with more than four strong interactions: in other words, if some hydrogen bonds are bifurcated. There are two different local geometries for the molecules that produce such a low energy state. In one configuration, the central water molecule shares one of its protons between two neighbour oxygens; in the other configuration, one of the two lone pairs accepts protons from two different molecules. In both cases, the central molecule has five neighbours instead of four. That such bifurcated bonds can occur is well known from hydrate crystals¹⁷. Our interpretation is supported by quantum mechanical calculations on water molecule clusters, which demonstrate that a symmetric¹⁸ or asymmetric¹⁹ energy-optimized bifurcated hydrogen bond is stronger than a conventional linear bond ($\delta E \approx -4 \text{ kJ mol}^{-1}$), and by experimental results²⁰ that can be interpreted in terms of the existence of four- and five-bonded water molecules²¹.

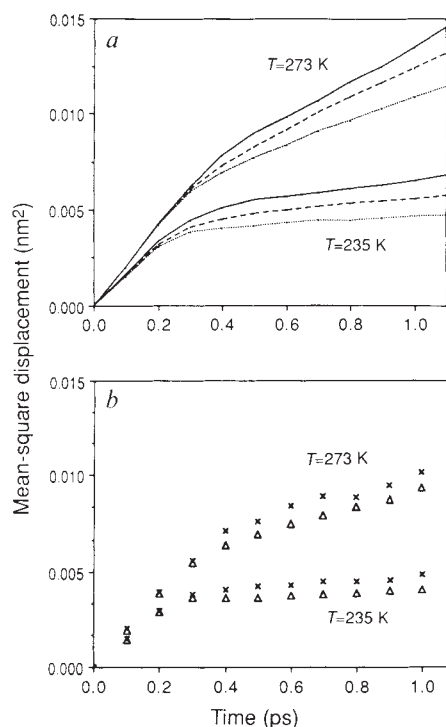
If the mobility of the liquid were related to the local density as expected from a picture of local free volume²², then one would expect the molecular mobility to increase as the density decreases. Instead, the opposite trend is observed¹⁴; Fig. 3 shows the density dependence of the self-diffusion coefficient, D , which we calculate from the mean square displacement. We find that lowering the density by 10% (from 1.0 to 0.9) decreases the diffusivity by a factor of two at 273 K, and by a factor of 10 at 235 K. We find a similar sensitivity to density for the times taken for the molecules to re-orient.

These results can be used to provide a microscopic support for the hypothesis that switching from one 'regular' hydrogen bond arrangement to another is facilitated by the local arrangement associated with higher local density. Accordingly, we relate structural data extracted from the quenched configurations to dynamical data extracted from the corresponding original molecular dynamics sequences. We calculate the mean square displacement $\langle r^2(t) \rangle_n$ for groups of molecules with numbers of neighbours $n = 4, 5$ and 6 (Fig. 4a). We observe a significantly higher mobility for molecules with more than four neighbours, confirming that molecular mobility is enhanced in the presence of 'extra' molecules in the first coordination shell.

To confirm that the higher mobility of the five-coordinated molecules compared with four-coordinated molecules is not simply a result of different bonding energies, we plot in Fig. 4b the mean square displacement for five- and four-coordinated molecules with the same E_i . We see that the five-coordinated molecules have a higher mobility than the four-coordinated molecules. Analogous results are obtained for rotational diffusion.

Our results support the possibility that the presence of an 'extra' molecule (in the first coordination shell) lowers the energy barriers for translational as well as rotational motions, and that the faster dynamics are not due to a weaker local bonding, but to a progressive reduction of the tetrahedral symmetry of the local minima in the potential-energy surface and the introduction of new local minima. Thus the extra molecule can 'catalyse' the restructuring of the hydrogen bond network when the thermal energy kT is much smaller than the hydrogen bond energy. An analogy with the influence of inert (hydrophobic) solutes is

FIG. 4 a, Mean square displacement $\langle r^2(t) \rangle_n$ for groups of molecules with number of neighbours n equal to four (—), five (---) and six (····) for $T = 235$ and 273 K. b, Mean square displacement $\langle r^2(t) \rangle_{n, E_i}$ for molecules with four (Δ) and five ($\times$) neighbours and binding energy E_i between -110 and -115 kJ mol^{-1} . The E_i bin between -110 and -115 kJ mol^{-1} has been selected because in this range one can find a sufficiently large number of both 5- and 4-coordinated molecules. As seen in Fig. 2, the average coordination number for this range of energies is ~ 4.5 . $\langle r^2(t) \rangle_n$ is calculated by following the time development of molecules for a time interval of 1.2 ps centred on a selected quenched configuration. The classification of the molecules according to the number of neighbours n is performed on the quenched configuration at the time $t = 0.6 \text{ ps}$. The average density here is 1.0 g cm^{-3} .



helpful. The decreased mobility and the increased structural order in hydrophobic hydration shells^{10,11} can be understood as consequences of the lower local density of water, which arise because the water is effectively locally diluted by the hydrophobic particle. The presence of the inert hydrophobic particle prevents the water molecules in the hydration shell being approached by more than four other molecules; this prevents the creation of low-activation-energy barriers and slows down the molecular mobility (this idea has received recent experimental support: see U. Kaatz and R. Pottel, preprint). The opposite effect—the creation of increased mobility by increasing the (global) density—is the well known ‘anomalous motional behaviour’ of water under compression^{8,9}. □

Received 20 July; accepted 26 September 1991.

- Angell, C. A. *Ann. Rev. Phys. Chem.* **34**, 593–611 (1983).
- Dore, J. & Teixeira, J. (eds) *Hydrogen Bonded Liquids* (Kluwer, Dordrecht, 1991).
- Stillinger, F. H. *Science* **209**, 451–457 (1980).
- Grünwald, E. *J. Am. chem. Soc.* **108**, 5719–5726 (1986).
- Narten, A. H. & Levy, H. A. *Science* **165**, 447–454 (1969).
- Giguère, P. A. *J. chem. Phys.* **87**, 4835–4839 (1987).
- Walrafen, G. E., Hokenmady, M. S., Yang, W.-H., Chu, Y. C. & Monosmith, B. *J. phys. Chem.* **93**, 2909–2917 (1989).
- Lang, E. W. & Lüdemann, H. D. in *High Pressure NMR* (ed. Jonas, J.) (Springer, Berlin, 1991).
- Pottel, R., Asselborn, E., Eck, R. & Tresp, V. *Ber. Buns. phys. Chem.* **93**, 676–688 (1989).
- Geiger, A., Rahman, A. & Stillinger, F. H. *J. chem. Phys.* **70**, 273–276 (1979).
- Huot, J.-Y. & Jolicoeur, C. in *The Chemical Physics of Solvation* (eds Dogonadze, R. R., Kalman, E., Kornyshev, A. A. & Ulstrup, J.) (Elsevier, Amsterdam, 1985).
- Angell, C. A. *Nature* **331**, 206–211 (1988).
- Green, J. L., Durben, D. J., Wolf, G. H. & Angell, C. A. *Science* **249**, 649–652 (1990).
- Geiger, A., Mausbach, P. & Schnitker, J. in *Water and Aqueous Solutions* (eds Neilson, G. W. & Enderby, J. E.) (Hilger, Bristol, 1986).
- Stillinger, F. H. & Weber, T. A. *J. phys. Chem.* **87**, 2833–2840 (1983).
- Weber, T. A. & Stillinger, F. H. *J. chem. Phys.* **87**, 3252–3253 (1987).
- Falk, M. & Knop, O. in *Water: A Comprehensive Treatise* (ed. Franks, F.) (Plenum, New York, 1972).
- Newton, M. D. *J. phys. Chem.* **87**, 4288–4292 (1983).
- Newton, M. D., Jeffrey, G. A. & Takagi, S. *J. Am. chem. Soc.* **101**, 1997–2010 (1979).
- Bosio, L., Chen, S. H. & Teixeira, J. *Phys. Rev. A* **27**, 1468 (1983).
- Sciortino, F., Geiger, A. & Stanley, H. E. *Phys. Rev. Lett.* **65**, 3452–3455 (1990).
- Cohen, M. H. & Turnbull, D. *J. chem. Phys.* **31**, 1164 (1959).
- Stillinger, F. H. & Rahman, A. *J. chem. Phys.* **60**, 1545–1557 (1974).
- Berendsen, H. J. C., Postma, J. P. M., von Gunsteren, W. F. & Hermans, J. in *Intermolecular Forces* (ed. Pullman, B.) 331 (Reidel, Dordrecht, 1981).
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. *J. chem. Phys.* **79**, 926 (1983).

ACKNOWLEDGEMENTS. We thank S. Grotzer, H. Larraide, P. H. Poole, S. Sastry, J. Teixeira and especially C. A. Angell. We also thank R. Baumert, M.-C. Bellissent-Funel, R. Bieshaar, F. Brüge, U. Essmann, G. G. Malenkov, V. Martorana, R. Pottel and D. Stauffer for helpful comments on the manuscript. This work was supported by BP, ONR, NSF, Fonds der Chemie and HLRZ Jülich.

Ocean circulation beneath the Ronne ice shelf

K. W. Nicholls, K. Makinson & A. V. Robinson

British Antarctic Survey, Natural Environment Research Council,
High Cross, Madingley Road, Cambridge CB3 0ET, UK

THE intimate thermal contact between the base of Antarctic ice shelves and the underlying ocean enables changes in climate to have a rapid impact on the outflow of ice from the interior of Antarctica^{1,2}. Furthermore, water modified by passage under ice shelves, particularly in the Weddell Sea, is believed to be an important constituent of Antarctic Bottom Water³—a water mass that can be observed as far north as 50° N in the deep oceans⁴. Antarctic Bottom Water is both cold and oxygen-rich, and plays an important part in the cooling and ventilation of the world's oceans. Because of the difficulty in gaining access, the oceanographic regime beneath ice shelves is very poorly sampled⁵. By successfully drilling through the ice, however, we were able to obtain oceanographic data from beneath the largest Antarctic ice shelf, the Ronne–Filchner ice shelf in the southern Weddell Sea. We find that our data agree well with the predictions of a relatively simple oceanographic plume model of sub-ice-shelf circulation⁶.

This model can therefore be used with some confidence to investigate the links between climate changes, ice-shelf melting and bottom-water production.

In January 1991, we used a pressurized hot-water drill to penetrate ice 562 m thick, 300 km from the northern limit of the Ronne ice shelf (Fig. 1). The drill was powered by three heat exchangers, providing a total heating power of 240 kW. A borehole pump, installed in an auxiliary borehole, retrieved enough meltwater to remove the need for continuous snow-melting during drilling. In total, 4.5 tonnes of aviation turbine fuel were burned in the heat exchangers, enabling a hole 20–25 cm in diameter to be drilled over a period of about a week, and then to be kept open for a further five days.

A borehole-deployable conductivity–temperature–depth (CTD) and water-sampling system was used to obtain several CTD profiles, and to retrieve five water samples from the ocean beneath the ice shelf (the samples are currently being analysed for isotope content). The accuracy of the temperature sensor on the CTD was 0.01 °C; the accuracy of the derived salinity was 0.015. Sensor calibrations were checked before and after the field work. The graphs in Fig. 2 show the potential temperature and salinity profiles that were obtained over a 17-h period. The upper 120 m of the water column is well mixed, with a potential temperature and salinity of ~ -2.30 °C and 34.53; the potential temperature and salinity over the next 130 m increase first slowly, then relatively rapidly to -2.07 °C and 34.64; from there to the bottom, at 850 dbar (85 bar), there is a slow increase in potential temperature and salinity to maxima of -2.02 °C and 34.65. To place the new data in context, the measured water characteristics have been plotted on a potential-temperature/salinity diagram (Fig. 3), together with the typical characteristics of the principal water masses of the Weddell Sea.

The data are consistent with the theory that the large-scale circulation under ice shelves is dominated by deep thermohaline convection. This theory has been advanced by Robin⁷, and expanded by several authors^{3,8,9}, most recently by Jenkins⁶. Western shelf water (WSW) is formed in the open lead north of the Ronne ice front by salt rejection from sea-ice production¹⁰, and has a temperature of ~ 1.9 °C. The theory proposes that it then drains down the sea-floor slope under the ice shelf (Fig. 4) and arrives unmodified at the base of the thick ice flowing off the continent. There the ice depth can be in excess of 1.5 km,

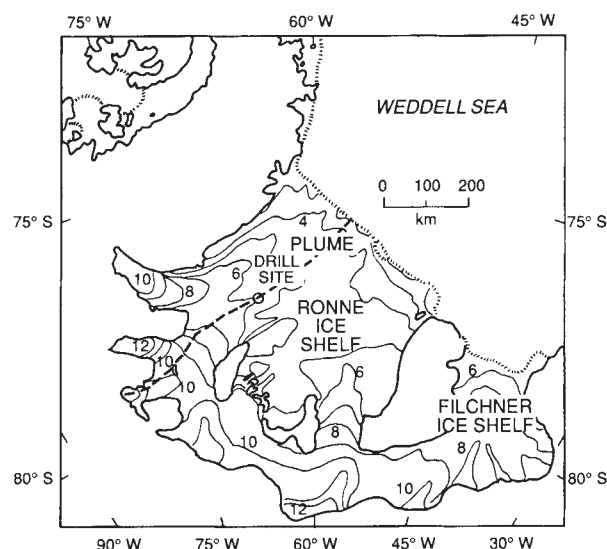


FIG. 1 Map of the Ronne ice shelf showing the location of the drill site (77° 36' S, 65° 28' W) and the path of the modelled plume. Contours give ice thickness in hundreds of metres¹⁴.

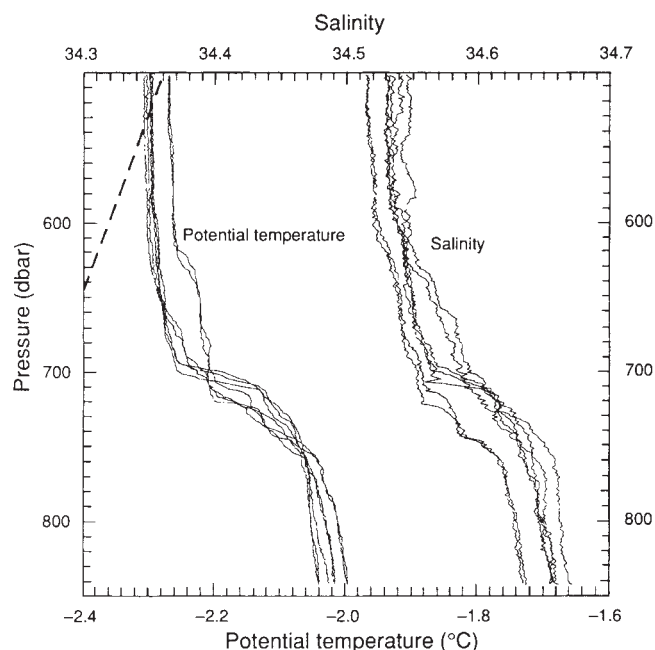


FIG. 2 Potential temperature and salinity profiles. The broken line is the *in situ* freezing point (corrected for potential temperature). The ice-shelf base is at 492 dbar. The variability in the mid-depth transition region is caused mainly by internal wave activity. The variability throughout the remainder of the water column is probably the effect of a tidal ellipse⁵ sweeping under the borehole water columns with slightly different characteristics.

with a corresponding pressure melting point below -3°C at its base. The relatively warm WSW causes rapid melting, and a buoyant mix of WSW and glacial meltwater ascends along the base of the thinning ice shelf, entraining more 'warm' WSW as it goes and thereby maintaining its capacity to continue melting. As the plume ascends, the increase in the local pressure melting point causes a gradual decrease in the melt rate. This, together with the decreasing basal slope of the ice shelf, reduces the buoyancy driving force; the plume slows, and the rate of entrainment of WSW from below decreases. A point finally arrives when the plume, cooled by the earlier melting at greater depths, becomes supercooled with respect to its surroundings. Frazil ice crystals then precipitate out and adhere to the base of the ice shelf, causing a layer of slightly saline ice. This process of redistribution of ice from greater depths to shallower areas has been described by Lewis and Perkin¹¹. Although the basal melting has stopped, the ice shelf continues to thin because of the horizontal spreading of the ice towards the seaward margin. The entrainment of the more saline WSW, and, to a minor degree, the salt rejection from frazil ice formation, increases the density of the plume until it finally separates from the base of the still-thinning ice shelf, ultimately emerging at the ice front as ice-shelf water (ISW) at mid-depth.

Jenkins⁶ has modelled the evolution of a plume to the point where it detaches from the base of the ice shelf, and has shown that the principal factors controlling its characteristics are the depth at the ice-shelf grounding line and the basal slope of the ice shelf. Here, the observed plume characteristics are compared with the predictions from the Jenkins model, which takes a plausible plume path from the grounding line to the ice front (Fig. 1), through the drill site.

A cold buoyant layer overriding a warmer, more saline layer, is clearly seen in the graphs in Fig. 2. The heavy broken line in Fig. 2 gives the melting point for the top few metres of the water

column, and shows that the water near the ice-shelf base is, for some of the time, supercooled. If the upper layer is the plume, it has clearly not yet detached from the ice shelf and is presumably precipitating frazil ice. Evidence for the build-up of frazil ice was provided by a borehole calliper which showed that, even after repeated drilling, the bottom 14 m of the hole could not be drilled out to give a smooth bore. This is thought to have been because of the presence of a layer of slush at the base of the ice shelf. A radio-echo sounder used at the drill site indicated an ice depth to the first reflector of 517 ± 2 m. Assuming this to be the thickness of fresh-water ice, there must be a basal layer of ~ 30 m of relatively consolidated saline ice underlain by the 14 m of slush. For this part of the plume, Jenkins' model⁶ predicts an accumulation rate of saline ice at the ice-shelf base of $\sim 0.3 \text{ m yr}^{-1}$. All these points imply that basal freezing conditions extend upstream for many kilometres. The extrapolation of a line through the temperature-salinity (T - S) data in Fig. 3 intercepts the WSW domain, strongly implicating WSW as one component in a simple mixing process. If the water column does indeed consist of WSW with well defined temperature and salinity, modified only by the melting of fresh-water ice, the T - S characteristic should be a straight line¹² with a gradient of $\sim 2.4^{\circ}\text{C}$. A line with that gradient is shown in Fig. 3, and corresponds well with the observations. The data show that the inflowing WSW, by the time it reaches the drill site at least 300 km from its source region, has already undergone considerable conversion by mixing with the ISW above, an interaction not included in the Jenkins' model (although another simulation of sub-ice-shelf circulation suggests that such an interaction is likely¹³). The model has been re-run using 'WSW' characteristics matching those observed at the drill site in the deeper water mass. The plume characteristics that result for the drill site location (A. Jenkins, personal communication) are indicated in

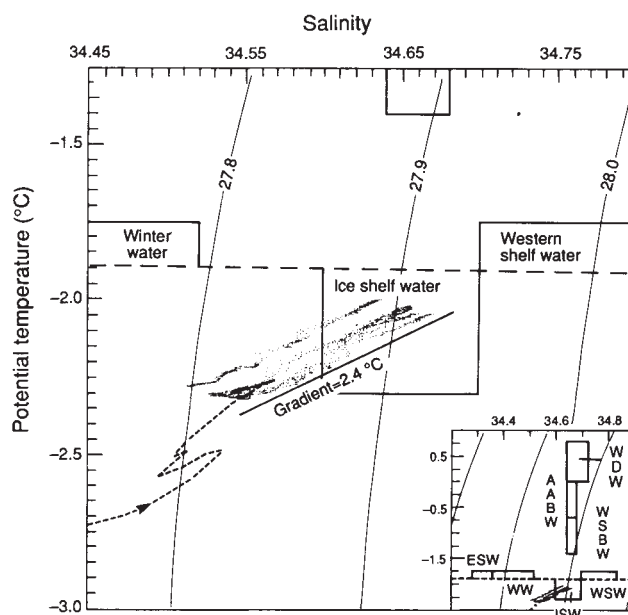


FIG. 3 Potential temperature/salinity diagram showing relationship between the data (dots) and the principal water types in the Weddell Sea¹⁵ (WDW, warm deep water; WW, winter water; WSW, western shelf water; ISW, ice-shelf water; WSBW, Weddell Sea bottom water; AABW, Antarctic Bottom Water; ESW, eastern shelf water). Fine lines, potential density isopleths. Broken line, trajectory of the temperature-salinity characteristic of the evolving plume in Jenkins' model⁶; straight broken line, freezing point the water would have at the sea surface. Inset, modelled plume characteristics at the drill site itself.

FIG. 4 Illustration of the deep thermohaline convection theory (see text). Tidal action¹⁶ and prevailing winds result in a transport of freshly formed sea ice away from the lead at the ice front, maintaining the exposure of the sea to the atmosphere and thereby intensifying sea-ice production and salt release¹⁰. Near the ice front, rapid basal melting is caused by 'warm' water being introduced from the open sea by tidal action. See text for further details.

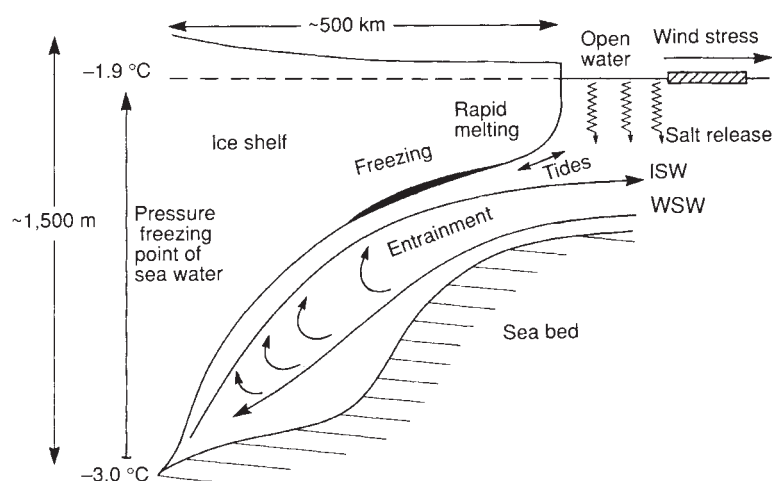


Fig. 3. The broken line shows the trajectory of the T - S characteristics of the modelled plume as it ascends the base of the ice shelf; the box shows the predicted characteristics at the drill site.

The potential temperature observed in the plume at the drill site was -2.27 to -2.31 °C, compared with -2.30 °C given by the model; close agreement was also found between the measured salinity, 34.515–34.555, and the model result, 34.55. The model is essentially one dimensional and therefore neglects Coriolis effects and any effects of topography (other than the basal topography of the ice shelf along the plume path). Tidal effects have also been neglected. Nevertheless, the agreement between such a simple model and the observed oceanography is impressive, lending support to the theory that WSW is the ultimate source water for ISW, and that thermohaline forcing is the principal driving for the large-scale circulation beneath ice shelves. WSW production is controlled by the climate of the Weddell Sea and, as ISW is thought to have an important role in bottom water formation, there exists a strong link between Antarctic climate and bottom water production. Those climatic conditions can similarly affect the basal melt rates of the ice shelf, and, therefore, the outflow of ice from the continent.

Direct oceanographic observations from other carefully selected locations on the ice shelf are needed to control future, larger-scale numerical models before a complete picture of the general circulation beneath the Ronne ice shelf can be formed, allowing the influence of sub-ice-shelf processes on the Antarctic oceanographic regime to be assessed. This will help provide the realistic southern boundary conditions essential for accurate global ocean and atmospheric modelling. □

Large-scale variation in lithospheric structure along and across the Kenya rift

KRISP Working Party*

THE Kenya rift is one of the classic examples of a continental rift zone: models for its evolution range from extension of the lithosphere by pure shear¹, through extension by simple shear², to diapiric upwelling of an asthenolith³. Following a pilot study in 1985⁴, the present work involved the shooting of three seismic refraction and wide-angle reflection profiles along the axis, across the margins, and on the northeastern flank of the rift (Fig. 1). These lines were intended to reconcile the different crustal thickness estimates for the northern and southern parts of the rift^{4–6} and to reveal the structure across the rift, including that beneath the flanks. The data, presented here, reveal significant lateral variations in structure both along and across the rift. The crust thins along the rift axis from 35 km in the south to 20 km in the north; there are abrupt changes in Moho depth and uppermost-mantle seismic velocity across the rift margins, and crustal thickening across the boundary between the Archaean craton and Pan-African orogenic belt immediately west of the rift. These results suggest that thickened crust may have controlled the rift's location, that there is a decrease in extension from north to south, and that the upper mantle immediately beneath the rift may contain reservoirs of magma generated at greater depth.

Combined seismic refraction and wide-angle reflection profiling yields reduced time record sections from which the relation of travel time and amplitude to distance can be determined for particular seismic phases travelling through the Earth. For the three profiles completed during the Kenya Rift International Seismic Project (KRISP90, January and February

Received 22 July; accepted 25 September 1991.

1. Budd, W. F., McInnes, J., Janssen, D. & Smith, I. N. in *Dynamics of the West Antarctic Ice Sheet* (eds. Van Der Veen, C. J. & Oerlemans, J.) (Reidel, Dordrecht, 1985).
2. Thomas, R. H. in *Glaciers, Ice Sheets, and Sea Level: Effects of a CO₂-Induced Climatic Change*. DOE/ER/60235-1 (National Research Council, US Department of Energy, Washington DC, 1985).
3. Foldvik, A. & Gammelsrød, T. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **67**, 3–17 (1988).
4. Dietrich, G., Kalle, K., Krauss, W. & Siedler, G. *General Oceanography*, 2nd ed. (Wiley, New York, 1980).
5. Foster, T. D. *J. geophys. Res.* **88**, 2556–2564 (1983).
6. Jenkins, A. *J. geophys. Res.* (in the press).
7. Robin, G. de Q. *J. Glaciol.* **24**, 259–271 (1979).
8. Hellmer, H. H. & Olbers, D. J. *Antarct. Sci.* **1**, 325–336 (1989).
9. MacAyeal, D. R. in *Oceanology of the Antarctic Continental Shelf, Antarctic Research Series 43*, (ed. Jacobs, S. S.) 133–143 (American Geophysical Union, Washington DC, 1985).
10. Gill, A. *Deep Sea Res.* **20**, 111–140 (1973).
11. Lewis, E. L. & Perkin, R. G. *J. geophys. Res.* **91**, 11756–11762 (1986).
12. Gade, H. G. *J. phys. Oceanogr.* **9**, 189–198 (1979).
13. Hellmer, H. H. & Olbers, D. J. *Antarct. Sci.* (in the press).
14. Pozdnyev, V. S. & Kurinin, R. G. *Antarkt. Dokl. Komissii*, **26**, 66–71 (1987).
15. Foldvik, A., Gammelsrød, T., Slotsvik, N. & Tørrisen, T. *Polar Res.* **3**, 209–226 (1985).
16. Gammelsrød, T. & Slotsvik, N. *Polarforschung* **51**, 101–111 (1981).

ACKNOWLEDGEMENTS. We thank B. Crawford for assistance during fieldwork and A. Jenkins for discussions.

* C. Prodehl, J. Mechie, W. Kaminski, K. Fuchs, C. Grosse, H. Hoffmann, R. Stangl, R. Stellrecht (Geophysikalisches Institut, University of Karlsruhe, Germany). M. A. Khan, P. K. H. Maguire, W. Kirk (Geology Department, University of Leicester, UK). G. R. Keller, A. Githui, M. Baker (Department of Geological Sciences, University of Texas, El Paso, USA). W. Mooney, E. Orley, J. Luetgert (US Geological Survey, Menlo Park, California). B. Jacob (Geophysics Section, DIAS, Ireland). H. Thybo (Institut für Almen Geologi, Københavns Universitet, Denmark). M. Demartin, S. Scarascia (CNR, Istituto Geofisica Litosfera, Milano, Italy). A. Hirn (IPG, Paris, France). J. R. Bowman (School of Earth Sciences, ANU, Canberra, Australia). I. Nyambok, S. Gaciri, J. Patel, E. Dindi (University of Nairobi, Kenya). D. H. Griffiths, R. F. King (Department of Geological Sciences, The University, Birmingham, UK). A. E. Mussett (Department of Geological Sciences, The University, Liverpool, UK). L. W. Braille (Department of Earth and Atmospheric Sciences, Purdue University, Indiana, USA). G. Thompson (Stanford University, Menlo Park, California, USA). K. Olsen (Los Alamos National Laboratory, New Mexico, USA). S. Harder (Texas A&M University, College Station, Texas, USA). R. Veas, D. Gajewski, A. Schulte (Institut für Geophysik, University of Clausthal, Germany). J. Obel (Survey of Kenya, Nairobi, Kenya). F. Mwango (Ministry of Water Development, Nairobi, Kenya). J. Mukinya (Department of Fisheries, Nairobi, Kenya). D. Riaroh (Ministry of Energy, Nairobi, Kenya).

1990), data quality was good (Fig. 2). We have interpreted the profiles mainly using two-dimensional ray tracing methods⁷. Perturbation of the models has shown that resolution of the velocity and depth to interface may be taken as better than 3% and 7% respectively, depending on the uniformity of the local structure and on the number of ray-paths intersecting the particular volume of the model.

The observed sediment and volcanic velocities (2.0 – 5.2 km s^{-1}) and thicknesses (up to 5.5 km) are equivalent to those found previously within the Kenya rift^{6,8}. The velocities in the upper crystalline crust, derived mainly from analysis of the P_g phase (Fig. 2), are between 5.9 and 6.2 km s^{-1} beneath all the lines, and are thus consistent with material of granitic composition. The velocity of the P_g refraction can be measured to better than 0.05 km s^{-1} . There is thus a significant change in velocity from 5.9 to 6.0 km s^{-1} across the Nandi Escarpment on line D (Fig. 3b). This marks the surface expression of the boundary between the Archaean Nyanza craton and the Pan-African Mozambique orogenic belt. Further east, beneath the rift, the velocity increases to 6.1 km s^{-1} . The extensive Cenozoic volcanism associated with the rift suggests this increase could be caused by the presence of basic dykes.

Analysis of the intracrustal phases $P_{11}P$ and P_{11} shows that there is a well defined velocity change to 6.4 – 6.6 km s^{-1} at ~ 10 – 15 km beneath each line, marking a change to a more basic average composition. This small velocity range probably masks considerable heterogeneity in both composition and physical

state. Heat flow values in Kenya⁹, although scattered, indicate an increase of 200 – 300°C across the rift margins at a depth of $\sim 15 \text{ km}$. Such a temperature increase would decrease the velocity by 0.1 – 0.2 km s^{-1} (refs 10–12). This velocity decrease could be compensated for by intrusion of more basic igneous material into the 6.4 – 6.6 km s^{-1} layer beneath the rift.

Examination of the topography of this mid-crustal boundary beneath Lake Bogoria, the only region for which well defined earthquake focal depths are available¹³, reveals that the cut-off depth of 12 km is definitely within the upper granitic layer, the brittle-ductile transition therefore being $\sim 3 \text{ km}$ above the layer of more basic composition.

A layer with velocity $\sim 6.8 \text{ km s}^{-1}$ is identified at the base of the crust beneath all three lines, except at the northern end of the axial profile. Even there, the model resolution does not preclude the presence of a layer up to 2 km thick with this velocity within the lowermost crust. The reflectivity of the upper surface of this layer varies considerably. Beneath the Nyanza craton to the west of the Nandi Escarpment, the top of this layer is defined by a sharp discontinuity identified from a high-amplitude reflection from the Lake Victoria shot (Fig. 2b). To the east of the rift on the same line, and also beneath the flank line, both of which lie on the Mozambique orogenic belt, the base of the crust is marked by an increasing velocity gradient. This difference suggests that the orogenic belt includes a systematic increase in mafic material towards the base of the crust, whereas the craton includes a well

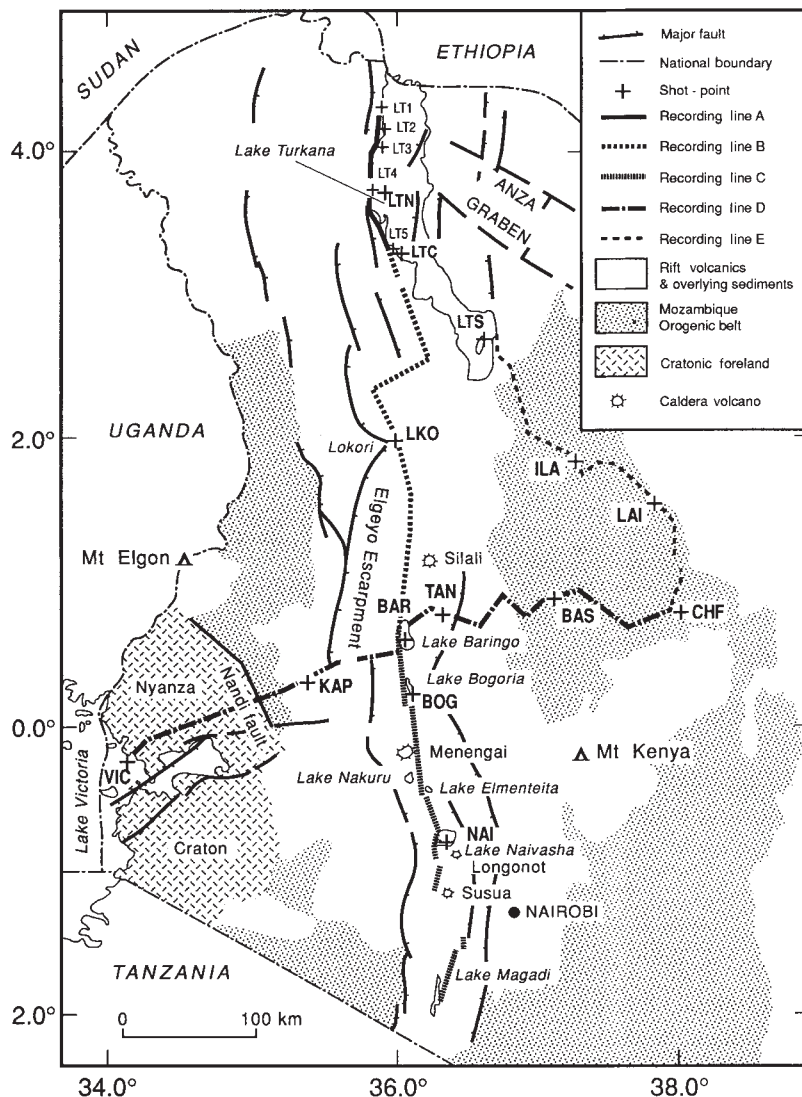


FIG. 1 KRISP 90 location map.

defined boundary separating a mafic layer at its base from the material above.

The 6.8 km s^{-1} layer beneath the southern part of the rift correlates well with the raised topography of the Kenya dome. Wide-angle reflections indicate that beneath this part of the rift the layer has a well defined upper boundary, unlike the remainder of the Mozambique belt. If we estimate the temperature increase across the rift margins to be $300\text{--}500^\circ\text{C}$ and assume that the material is the same as that observed elsewhere within the orogenic belt, but intruded by Cenozoic mafic material, we obtain an equivalent velocity (assuming a standard continental geotherm) of $\sim 7.0\text{--}7.1 \text{ km s}^{-1}$ (ref. 12). This is somewhat lower than values observed for deep crustal intrusive material beneath older Phanerozoic continental rifts¹⁴ and passive margins¹⁵. The difference may be due to the amount of intrusion being less beneath the Kenya rift.

Anomalous upper mantle, characterized by a low P_n velocity (Fig. 3a) of $7.5\text{--}7.7 \text{ km s}^{-1}$, observed along the complete length

of the axial profile, is confined to the region beneath the rift itself. It abuts normal upper mantle with a P_n velocity of $8.0\text{--}8.1 \text{ km s}^{-1}$ beneath the rift flanks (Fig. 3b). From heat flow values⁹ and without wholesale melting of the lower crust, the temperature at the base of the crust beneath the rift cannot be more than $\sim 1,000^\circ\text{C}$. At this temperature the velocity of peridotitic mantle beneath the eastern flank of the rift (R. Altherr, personal communication) would be $\sim 7.7 \text{ km s}^{-1}$. If sub-Moho velocities as low as 7.5 km s^{-1} exist, then this layer beneath the Rift may contain pyroxene-enriched, olivine-depleted material and/or molten magma generated at greater depths. The strong wide-angle reflection, d_1 , from shot point LTN from a depth of 45 km beneath the northern part of the axial line (Fig. 3a) indicates that there are significant changes in the internal velocity structure of the anomalous uppermost mantle beneath the rift.

The cross-section along line D (Fig. 3b) shows considerable variation in crustal thickness. Beneath the western flank, crustal thickness varies from 37 km to 40 km, with a modest thickening

FIG. 2 a, Trace-normalized, low-pass filtered, reduced time-distance record section with phase correlations for shot-point LTN recorded to the south along the axial profile. Reduction velocity 6 km s^{-1} . Phase notation: a, P_g (upper crustal refraction); b, $P_{13}P$ (intracrustal reflection); c, P_mP (Moho reflection); d, P_n (uppermost mantle refraction); d_1 (uppermost mantle reflection). b, Trace-normalized, low-pass filtered, reduced time-distance record section with phase correlations for shot-point VIC recorded into line D. Phase notation: as for a, and $b'=P_{13}$ (intracrustal refraction); $b_1=P_{12}P$ (intracrustal reflection). The large range of amplitudes means that, at this scale, the weaker first arrivals (such as phase a) are difficult to identify.

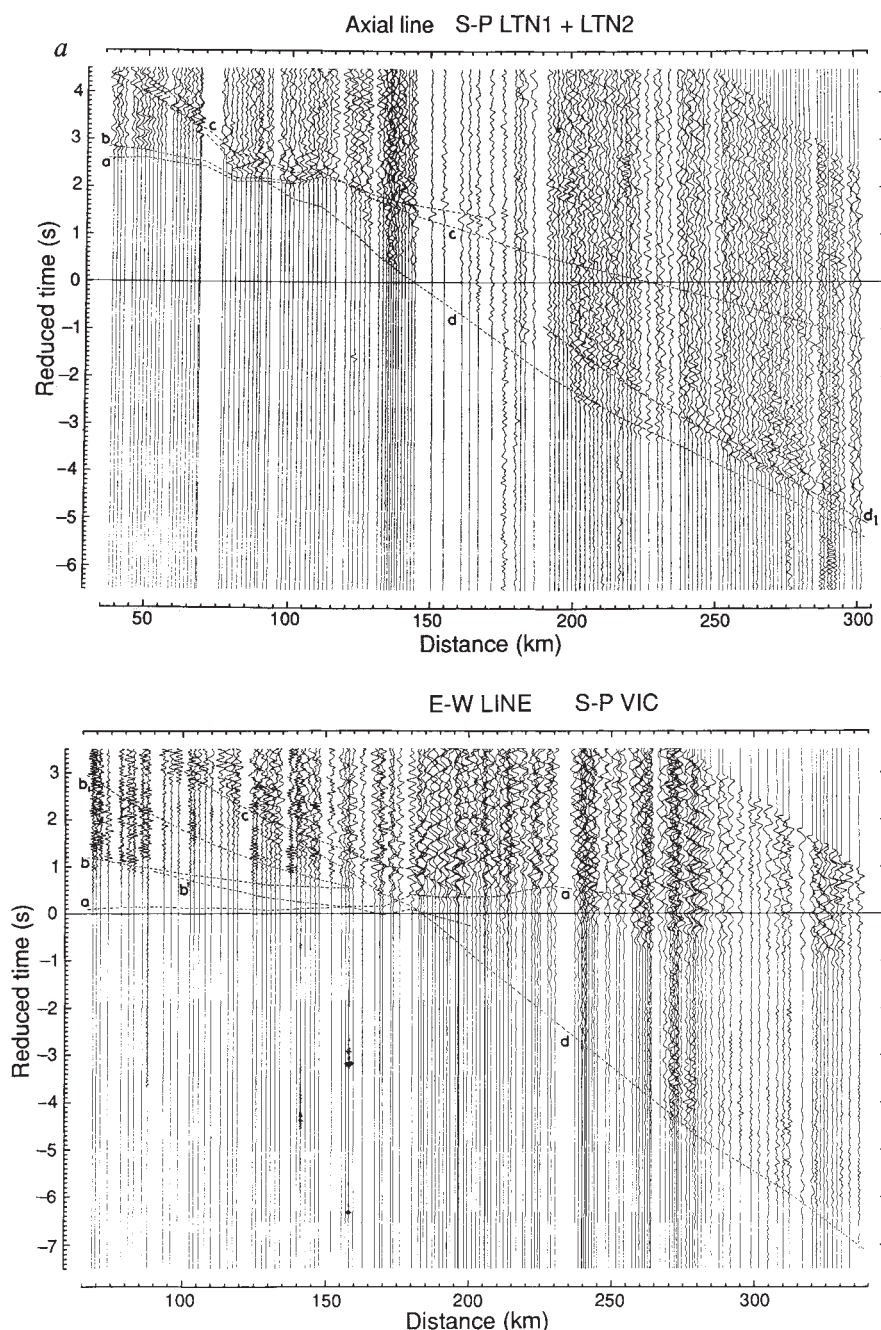
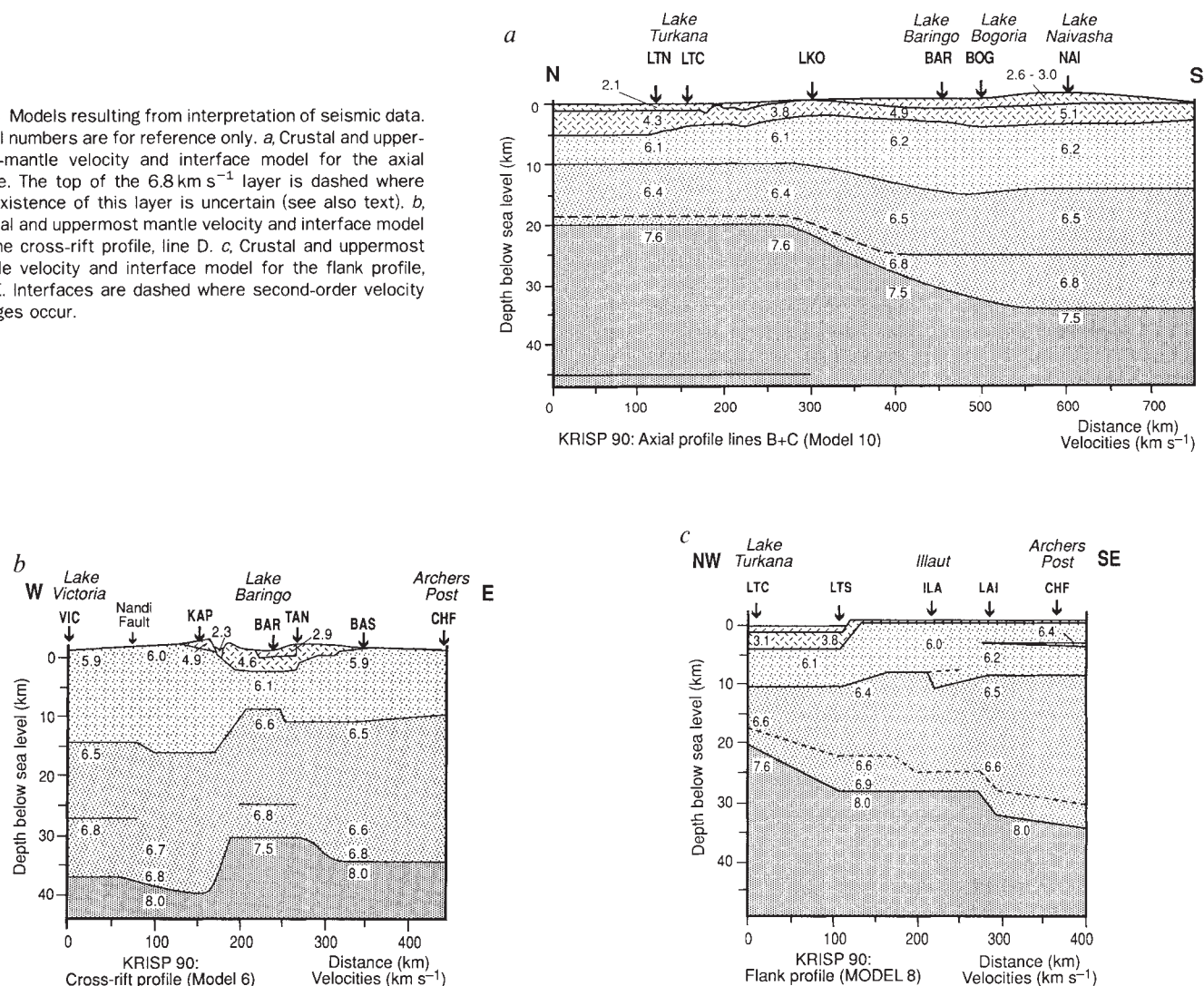


FIG. 3 Models resulting from interpretation of seismic data. Model numbers are for reference only. *a*, Crustal and uppermost-mantle velocity and interface model for the axial profile. The top of the 6.8 km s^{-1} layer is dashed where the existence of this layer is uncertain (see also text). *b*, Crustal and uppermost mantle velocity and interface model for the cross-rift profile, line D. *c*, Crustal and uppermost mantle velocity and interface model for the flank profile, line E. Interfaces are dashed where second-order velocity changes occur.



beneath that part of the Mozambique orogenic belt west of the rift. At the rift's western boundary near the Elgeyo escarpment, the crust thins considerably to 30 km below the rift axis at Lake Baringo. It thickens again to 35 km beneath the eastern flank at the same latitude. The crustal cross-section across the rift in this region is thus asymmetric.

The transition between the orogenic belt and the Nyanza craton is evident at all depths within the crust and seems to be abrupt and sub-vertical. The fact that the Cenozoic Rift occurs close to this transition could be due to the thickened Pan-African crust in this region creating a weak lithosphere with a higher proportion of more ductile (weaker) minerals such as quartz and feldspar, dominant in the crust, and a lower proportion of less ductile (stronger) minerals such as olivine, dominant in the upper mantle^{16,17}. Further weakening due to the raising of the lithospheric geotherm would accelerate the rifting process.

The model demonstrates that crustal thinning and anomalous mantle are essentially restricted to beneath the rift itself (Fig. 3*b*). This is consistent with results from recent teleseismic studies¹⁸ showing anomalous low-velocity material being constrained within the rift boundaries to depths of 200 km. It is not consistent with models² indicating that the major lithospheric thinning is offset beneath one of the rift flanks.

The model (Fig. 3*a*) supports the previous estimate⁵ for crustal thickness of 18.5 km beneath Lake Turkana. We have extended the model considerably, however, to show crustal thinning from 35 km beneath the apex of the Kenya dome to 20 km beneath Lake Turkana. Because the flank line lies between the Jurassic

Anza graben and the Tertiary Kenya rift to the south of Lokori (Fig. 1), we can state that this thinning along the Kenya rift with respect to that beneath the flank line must have been caused by the Tertiary rifting episode.

The main 10–15 km variation in crustal thickness along the axial profile can be correlated with topographic relief along the rift axis and with rift width. The crustal thinning to the north, together with the apparent splaying of the rifted region to the north of Lake Baringo, suggests a change to highly extended terrain, consistent with extension estimates of ~35–40 km in the northern Kenya rift¹⁹ as opposed to ~10 km over the apex of the Kenya dome¹. The broad zone of extension in the north, overlying thin crust, suggests a response to a passive lithospheric stretching event. The narrow zone of extension over the apex of the Kenya dome overlying thick crust, which in turn overlies an upper-mantle low-velocity zone¹⁸ extending down to 200 km depth, suggests less a response to stretching and more a response to an upward thermal perturbation of the lithosphere–asthenosphere boundary with thinning of the lithosphere from below. □

Received 8 April; accepted 29 October 1991.

1. Baker, B. H. & Wohlenberg, J. *Nature* **229**, 538–542 (1971).
2. Bosworth, W. *Geology* **15**, 397–400 (1987).
3. Mohr, P. *Eos. Trans. Am. Geophys. Un.* **68**, 721–736 (1987).
4. KRISP Working Group *Nature* **325**, 239–242 (1987).
5. Griffiths, D. H., King, R. F., Khan, M. A. & Blundell, D. J. *Nature* **229**, 69–71 (1971).
6. Henry, W. J. *et al. Geophys. J. Int.* **100**, 107–130 (1990).

7. Červený, V., Molotkov, I. A. & Pšenčík, I. In *Ray Method in Seismology* (University of Karlova, Prague, 1977).
8. Swain, C. J., Khan, M. A., Wilton, T. J., Maguire, P. K. H. & Griffiths, D. H. *J. geol. Soc. Lond.* **138**, 93–102 (1981).
9. Morgan, P. In *Continental and Oceanic Rifts* (ed. Palmason, G.) *Am. Geophys. Un. Geodyn. Ser.* **8**, 107–122 (1982).
10. Birch, F. *J. geophys. Res.* **65**, 1083–1102 (1960).
11. Christensen, N. I. *J. geophys. Res.* **70**, 6147–6164 (1965).
12. Christensen, N. I. *J. geophys. Res.* **84**, 6849–6857 (1979).
13. Young, P. A. V., Maguire, P. K. H., Evans, J. R. & Lafoley, N. d'A. *Geophys. J. Int.* **105**, 665–674 (1991).
14. Mooney, W. D., Andrews, M. C., Ginzburg, A., Peters, D. A. & Hamilton, R. M. *Tectonophysics* **74**, 327–348 (1983).
15. White, R. S. *et al. Nature* **330**, 439–444 (1987).
16. Vink, G. E., Morgan, W. J. & Zhao, W.-L. *J. geophys. Res.* **89**, 10072–10076 (1984).
17. Steckler, M. S. & ten Brink, U. S. *Earth planet. Sci. Lett.* **79**, 120–132 (1986).
18. Dalheim, H.-A., Davis, P. & Achauer, U. *J. Afr. Earth Sci.* **8**, 461–470 (1989).
19. Morley, C. K. *et al. J. Geol. Soc.* (in the press).

ACKNOWLEDGEMENTS. The following participated in the programme: D. Abuuru, B. Aichroth, T. Blake, J. Borongo, P. Bowry, P. Burton, N. Chroston, W. Clement, J. Cotton, P. Coward, T. Elvers, U. Enderle, F. Fischer, J. Geppert, M. Goldman, A. Haston, C. Horan, S. Hughes, E. Imana, J. Jepsen, K. Joehnk, R. Kaderabek, U. Kaestner, N. Kamundia, J. Karanja, J. Kingeo, W. Lutter, H. Macharia, M. Maistrello, J. Mansfield, J. Mburu, D. Mutai, J. Ndede, J. Ndombi, J. Oberbeck, B. O'Neill, B. O'Reilly, W. Okoth, T. Parsons, P. Readman, D. Roberts, B. Rippere, A. Rueger, F. Schjøth, M. Schoch, W. Teasdale, J. Tongue, F. Vine, C. Wafula, G. Wallace, R. Young, D. Zola. We thank the Kenya Government and the officials who made the work possible. The University of Nairobi, Egerton College, the Department of Mines and Geology, the Ministry of Water Development, the Survey of Kenya, Contratours, Inside Africa Safaris, Aquadrill, Twiga Chemical Industries all contributed to the programme. The work was financed by the EEC under SCI contract 00064, the DFG in Germany, NSF in the US, and the NERC in the UK.

Floral colour changes as cues for pollinators

Martha R. Weiss

Department of Integrative Biology, University of California, Berkeley, California 94720, USA

PLANTS have evolved traits that enable them to influence directly the behaviour and movement of their pollinators. Here I show that flowers in at least 74 diverse angiosperm families undergo dramatic, often localized, colour changes which direct the movements of a variety of pollinators to the benefit of both participants. Floral colour change was first noted almost 200 years ago¹ and is known in a variety of species^{2–17}, but the prevalence and significance of the phenomenon have gone largely unrecognized. I find that retention of older flowers increases a plant's attractiveness to pollinators from a distance, that pollinators discriminate between floral colour phases at close range, and that the discrimination involves learning. The phenomenon of floral colour change is taxonomically widespread, morphologically variable (Fig. 1), and physiologically diverse. It has evolved independently in the angiosperms many times and provides a striking example of functional convergence.

In 1877 Charles Darwin forwarded to *Nature* a letter from his colleague, the naturalist Fritz Müller, in which Müller remarked on a multicoloured *Lantana* growing in the Brazilian forest³,

"We have here a *Lantana* the flowers of which last three days, being yellow on the first, orange on the second, purple on the third day. This plant is visited by various butterflies. As far as I have seen the purple flowers are never touched. Some species inserted their proboscis both into yellow and into orange flowers; others, as far as I have hitherto observed, exclusively into the yellow flowers of the first day. This is, I think, a rather interesting case. If the flowers fell off at the end of the first day the inflorescence would be much less conspicuous; if they did not change their colour much time would be lost by the butterflies inserting their proboscis in already fertilized flowers."

I have used *Lantana camara* flowers and two species of nymphalid butterflies, *Agraulis vanillae* and *Junonia coenia*, to examine Müller's ideas. *L. camara* flowers are yellow only on the day they open, turning orange and then red on subsequent days. Red flowers are maintained on the plant for 1 to 9 days, depending on the *Lantana* variety and climatic conditions. Only yellow flowers, which comprise between 9 and 33% of the total

TABLE 1 Butterfly visits to *L. camara* displays of varying size and reward level

						
	Large red	vs Large YOR	Small yellow	vs Large YOR	Small yellow	vs Large red
Fraction of trials in which display type received most visits	0.5*/7	6.5*/7	0/7	7/7	1/6	5/6
Total butterfly visits	132	264	137	270	155	205
χ^2 (1 d.f.)	44.00†		43.46†		6.94‡	

Butterflies chose to visit *Lantana camara* inflorescences on the basis of the size of floral display, rather than the level of nectar reward it offered. In 3 replicated sets of trials, *A. vanillae* butterflies were offered different pairs of floral displays which provided various combinations of size and reward level. 'Large yellow-orange-red' (YOR) inflorescences were unmodified, and contained 3–5 yellow, 3–5 orange, and 24–30 red flowers. Flowers were removed from inflorescences as necessary to produce the 2 other types of displays: 'small yellow' consisted of 3–5 rewarding yellow flowers, and 'large red' consisted of 25–35 non-rewarding red flowers. Butterflies were offered large red versus large YOR (7 replicates); small yellow vs large YOR (7 replicates); small yellow vs large red (6 replicates). Inflorescences were picked in the morning before the yellow flowers opened, and were kept in water. In each trial, 3 samples of each of the 2 offerings were placed inside a 2-m³ field cage, arranged alternately in water-filled jars, 0.5 m apart, in a hexagonal pattern. Ten to fifteen butterflies with previous experience in feeding on *L. camara* were present in the field cage during each half-hour trial. A 'visit' was counted when a butterfly landed on a inflorescence. Placement of each type of offering was rotated 60° between replicates of trials to control for possible position effects. A control trial, in which 6 identical samples were offered, showed no significant differences between positions (*G* test for goodness of fit, *G*=1.07, 1 d.f., n.s.). For each of the 3 sets of trials, I determined the fraction of trials in which a particular display type received more than 50% of the butterfly visits. Additionally, visits to each of the 3 samples of a given display type were grouped, and the total number of visits to each of the 2 offerings was compared using a χ^2 test for goodness of fit²⁷, with a null hypothesis of equal attractiveness.

* Value of 0.5 indicates that the outcome of the trial was tied, that is, both offerings received equal numbers of visits. † *P* < 0.001; ‡ *P* < 0.01.

floral display, offer nectar and pollen and are receptive (M.R.W., manuscript in preparation).

To determine whether large displays are more attractive to pollinators at a distance, I conducted three sets of choice tests inside a field cage, offering experienced butterflies natural or manipulated *Lantana* inflorescences which provided various combinations of display size and nectar reward. The trials included: (1) large displays offering a nectar reward equal to that of small displays; (2) two equally large displays offering different amounts of reward; and (3) large displays offering less reward than small displays. When display sizes differed, butterflies chose to visit the larger offering, regardless of its level of reward; when display sizes were equal, the more rewarding offering was preferred. These results were statistically significant (Table 1). Maintenance of older flowers on an inflorescence thus increases the plant's attractiveness to pollinators at a distance.

To test whether pollinators discriminate between floral colour phases at close range, I presented caged butterflies, experienced in feeding on *Lantana*, with yellow and red *Lantana* inflorescences. These insects discriminated between the colour phases, choosing to feed at the yellow, pre-change flowers significantly more often than would be expected from the abundance of the flowers (Table 2a).

The discrimination between *Lantana* colour phases apparently involves associative learning, at least in some insects.

TABLE 2 Insect visits to colour-changing flowers

Plant (family)	Insect or insect family (<i>n</i> insects or feeding bouts)	Mean % pre-change flowers	Total visits to pre-change flowers	Total visits to post-change flowers	χ^2 (1 d.f.)
(a) Visits of <i>Agraulis vanillae</i> and <i>Junonia coenia</i> to <i>L. camara</i> (caged)					
<i>Lantana camara</i> (Verbenaceae)	<i>J. coenia</i> (15) Nymphalidae, Lepidoptera	10.5	347	227	1524.13*
<i>Lantana camara</i> (Verbenaceae)	<i>A. vanillae</i> (43) Nymphalidae, Lepidoptera	13.7	436	435	973.77*
(b) Insect visits to colour-changing flowers (field)					
<i>Phyla nodiflora</i> (Verbenaceae)	Apidae (3) Hymenoptera	24	100	12	261.71*
<i>Phyla nodiflora</i> (Verbenaceae)	Hesperiidae (1) Lepidoptera	24	28	1	83.69*
<i>Phyla nodiflora</i> (Verbenaceae)	Calliphoridae (1) Diptera	24	20	1	58.43*
<i>Heliotropium anchusifolium</i> (Boraginaceae)	Apidae (8) Hymenoptera	46.5	346	29	315.72*
<i>Androsace lanuginosa</i> (Primulaceae)	Bombyliidae (1) Diptera	42	19	0	26.24*
<i>Androsace lanuginosa</i> (Primulaceae)	Calliphoridae (1) Diptera	42	14	1	16.22*
<i>Lobularia maritima</i> (Cruciferae)	Syrphidae (1) Diptera	7.1	14	1	*
<i>Lobularia maritima</i> (Cruciferae)	Tachinidae (1) Diptera	7.1	23	0	*
<i>Lobularia maritima</i> (Cruciferae)	Calliphoridae (2) Diptera	7.1	57	0	*
<i>Lupinus albus</i> (Leguminosae)	Apidae (3) Hymenoptera	38.5	96	8	127.17*

A wide variety of insects discriminate between floral colour phases, visiting pre-change flowers significantly more than would be expected based on the abundances of these flowers in inflorescences. Visits were counted when the insect probed at the flower and attempted to reach the pollen or nectar. The proportion of flowers in each colour phase was determined, and the distribution of the insects' visits was tested against expected values based on abundances of floral colour phases, using a Chi Square test for Goodness of Fit. In cases where the expected value was less than 5, a binomial expansion was used²⁷. a, Visitation patterns of two species of nymphalid butterflies, *A. vanillae* and *J. coenia*, on *L. camara*; these insects were observed in a flight cage. b, Visitation patterns of a variety of insects on a variety of plants, observed in the field, and grouped by family.

* $P < 0.001$.

Newly emerged naive butterflies visit both yellow and red flowers in their initial feeding bouts, but quickly come to concentrate their visits on the rewarding yellow flowers. The plant essentially 'teaches' the insect to focus its attention on sexually viable and rewarding flowers (Table 3).

In a broad survey of flowering plants, I found that colour-changing species were taxonomically and geographically diverse, occurring in at least 214 genera, 74 families and 33 orders worldwide. Conservatively then, colour change occurs in at least 21% of animal-pollinated angiosperm families.

Floral colour change benefits both plant and pollinator. The colour phase of a flower has been shown in several cases to provide an accurate indication of its sexual viability and reward status⁹⁻¹⁷. Comparing nectar volume, stigmatic condition and pollen availability in pre- and post-change flowers of 26 locally available species, I found that pre-change flowers offered nectar rewards in all 26 cases, whereas post-change flowers contained little or no nectar. In addition, pollen was available and stigmas appeared fresh and receptive in the pre-change flowers, but post-change flowers lacked pollen and appeared non-receptive. Thus plants receive efficient pollination service while pollinators are accurately directed to rewarding flowers.

A diverse array of pollinators discriminate between floral colour phases on a variety of colour-changing plants, as butterflies do on *Lantana*. I observed dipteran, hymenopteran and lepidopteran pollinators, representing 15 insect families, foraging on 28 different colour-changing plant species in the field. The insects consistently concentrated their visits on pre-change flowers, visiting them significantly more often than would be expected from the flowers' abundances (Table 2b). These insects were not necessarily the plants' natural pollinators: individuals of introduced species discriminated accurately, as did oppor-

tunistic nectar or pollen robbers. Thus colour-changing flowers and their pollinators need not have closely coevolved. Rather, the generality and power of the system depend on the plants' plasticity and the insects' ability to learn.

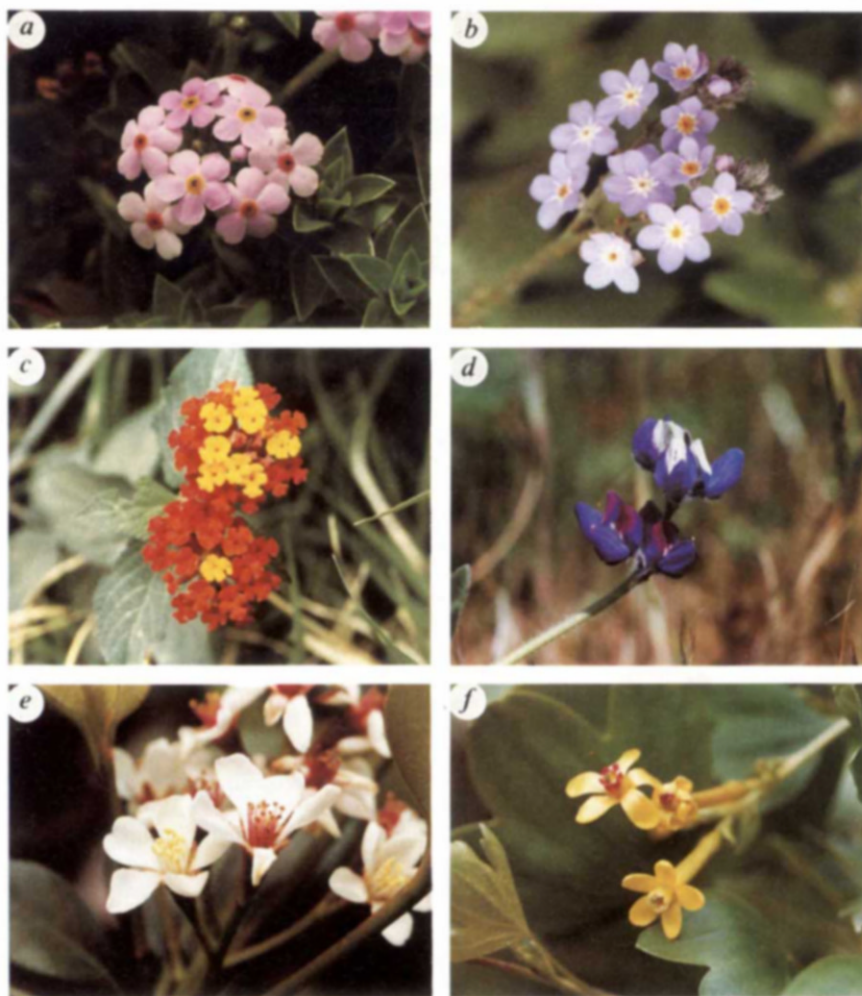
It is often assumed that plants have a largely passive role in plant-animal relationships. But plants are able, through a variety

TABLE 3 Mean preference indices for naive and experienced butterflies

	<i>Junonia coenia</i>	<i>Agraulis vanillae</i>
Naive	$\bar{\chi} = 5.8$ $s_{\chi} = 4.9$ $n = 5$	$\bar{\chi} = 3.3$ $s_{\chi} = 1.0$ $n = 8$
Experienced	$\bar{\chi} = 21.6$ $s_{\chi} = 26.3$ $n = 15$	$\bar{\chi} = 30.6$ $s_{\chi} = 34.0$ $n = 15$

Naive butterflies offered bicoloured *Lantana camara* inflorescences chose relatively fewer yellow and relatively more red flowers than did experienced individuals. Naive butterflies, which were not exposed to any flowers before they were offered *L. camara* inflorescences, were tested only once. Experienced butterflies, which had at least 2 days prior experience feeding on *L. camara*, were offered flowers on one or several occasions. For each butterfly, the colour of every flower probed was recorded; to standardize differences in abundances of floral colour phases, a preference index (PI) was calculated as: (number of visits to yellow flowers/number of yellow flowers)/(number of visits to red flowers/number of red flowers). A value of 1 indicates visitation proportional to abundances of colour phases; values above 1 indicate preferential visitation to yellow flowers. For experienced butterflies whose feeding preferences were observed on more than one occasion, mean PI values were used in calculating the overall mean. Mean PIs for naive and experienced butterflies were significantly different within each species: Mann-Whitney *U* test for *Junonia*, 66.5, $P < 0.01$; for *Agraulis*, 120, $P < 0.001$ (ref. 27). *n*, Number of butterflies.

FIG. 1 Examples of floral colour change. Floral colour changes may involve different parts of flowers as well as different physiological processes¹⁸. Floral parts which change include the corona or eye, corolla throat, nectar guide, banner petal spot, specialized petal, hypanthium, filaments, ovary or whole flower. Physiological processes may involve either appearance or loss of anthocyanin or carotenoid pigments. Appearance of anthocyanin is the most common mechanism of change (M.R.W., manuscript in preparation). In all cases, regardless of the flower part or the physiological processes involved in the colour change, the end result is the same: young, sexually viable and rewarding flowers are visually distinguishable from older flowers. *a*, *Androsace lanuginosa* Wall. (Primulaceae): eye changes from yellow through orange to red; *b*, *Myosotis* sp. (Boraginaceae): corona changes from yellow to white; *c*, *Lantana camara* L. (Verbenaceae): whole flower changes from yellow to red; *d*, *Lupinus nanus* Dougl. (Leguminosae): banner petal spot changes from white to purple; *e*, *Raphiolepis umbellata* Makino (Rosaceae): filaments change from white to red; *f*, *Ribes odoratum* Wendl. (Saxifragaceae): petals change from yellow to red.



of signals, to influence directly the behaviour and movement of the animals on which they depend. Colour signals may be used to indicate to birds which fruits are ripe and ready for dispersal^{19–21}; floral nectar volumes can influence the length of time a pollinator will stay on a given plant^{22,23}; specific floral odours

can attract the right pollinators at appropriate times^{24–26}. I have shown that plants in at least 74 different families use colour changes to direct their pollinators to rewarding and sexually viable flowers. Through such signals, plants are able to play a surprisingly active part in their interactions with animals. □

Received 7 August; accepted 17 September 1991.

1. Sprengel, C. K. *Das Entdeckte Geheimniss der Natur im Bau und in der Befruchtung der Blumen* (Vieweg, 1793).
2. Delpino, F. *Atti Soc. ital. Sci. nat.* **16**, 151–349 (1873).
3. Müller, F. *Nature* **17**, 78–79 (1877).
4. Müller, H. *Nature* **28**, 81 (1883).
5. Kugler, H. *Ergebn. Biol.* **19**, 143–323 (1943).
6. Vogel, S. *Öst. bot. Z.* **97**, 44–100 (1950).
7. Dunn, D. B. *Am. Midl. Nat.* **555**, 443–472 (1956).
8. Wainwright, C. M. *Bull. Torrey Bot. Club* **105**, 24–38 (1978).
9. Schaal, B. A. & Leverich, W. J. *S. West. Nat.* **25**, 280–282 (1980).
10. Gori, D. F. in *Handbook of Experimental Pollination Biology* (eds Jones, C. E. & Little, R. J.) (Van Nostrand Reinhold, New York, 1983).
11. Casper, B. B. & La Pine, T. R. *Evolution* **38**, 128–141 (1984).
12. Lamont, B. *Bot. J. Linn. Soc.* **90**, 145–155 (1985).
13. Mathur, G. & Mohan Ram, H. Y. *Phytomorphology* **36**, 79–100 (1986).
14. Eiskowitch, D. & Rotem, R. *Bot. Gaz.* **148**, 175–179 (1987).
15. Cruzan, M. B., Neal, P. R. & Willson, M. F. *Evolution* **42**, 505–515 (1988).
16. Gori, D. F. *Evolution* **43**, 870–881 (1989).
17. Delph, L. F. & Lively, C. M. *Evolution* **43**, 1252–1262 (1989).

18. Weiss, M. R. *Acta Hort.* **288**, 294–298 (1991).
19. Willson, M. F. & Whalen, C. J. *Evolution* **136**, 790–809 (1990).
20. Wheelwright, N. T. & Janson, C. H. *Am. Nat.* **126**, 777–799 (1985).
21. Greig-Smith, P. W. *Am. Nat.* **127**, 246–251 (1986).
22. Zimmerman, M. in *Plant Reproductive Ecology—Patterns and Strategies* (eds Lovett Doust, J. & Lovett Doust, L.) (Oxford University Press, 1988).
23. Pflumm, W. *Oecologia* **66**, 207–210 (1985).
24. Overland, L. *Am. J. Bot.* **47**, 378–382 (1960).
25. Robacker, D. C., Meeuse, B. J. D. & Erickson, E. H. *Bioscience* **38**, 390–398 (1988).
26. Williams, N. H. in *Handbook of Experimental Pollination Biology* (eds Jones, C. E. & Little, R. J.) (Van Nostrand Reinhold, New York, 1983).
27. Sokal, R. R. & Rohlf, F. J. *Biometry* (Freeman, New York, 1981).

ACKNOWLEDGEMENTS. I thank H. G. Baker and the late I. Baker, and E. A. Bernays, E. A. Dean, H. E. M. Dobson, S. Farmer, D. Greene, D. Harder, S. Mayer, J. P. Rosenthal, E. Smith, T. B. Smith, J. D. Villa, B. J. D. Meeuse and my family, for helpful comments; the many people who helped with field observations and brought colour-changing flowers to my attention; P. T. Spowart for the illustrations and assembling the plate, and R. Zuparko for identifying the flies. This research was supported by grants from the Arco Foundation, NSF, the J. S. Noyes Foundation, the Hardman Foundation, Soroptimists International, and the Association for Women in Science.

A *Drosophila* mutant defective in extracellular calcium-dependent photoreceptor deactivation and rapid desensitization

Rama Ranganathan*, Greg L. Harris†, Charles F. Stevens*‡§ & Charles S. Zuker*§¶

* Howard Hughes Medical Institute and Departments of Biology and § Neurosciences, University of California, San Diego, La Jolla, California 92093, USA

† Department of Biology and Molecular Biology Institute, San Diego State University, San Diego, California 92182, USA

‡ Salk Institute for Biological Studies, La Jolla, California 92093, USA

CALCIUM is involved in the adaptation of vertebrate photoreceptors to light^{1,2} and may have a similar role in invertebrate phototransduction^{3,4}. But the molecular mechanisms mediating this stimulus-dependent regulation are not well understood in any G protein-coupled transduction system. We have developed a preparation of isolated *Drosophila* photoreceptors that has allowed us to carry out an electrophysiological characterization of the light-activated response in these sensory neurons using patch-clamp techniques. We report here that extracellular calcium entering through the light-activated conductance is a key regulator of both the activation and deactivation phases of the phototransduction cascade, and that *inaC* mutant photoreceptors⁵ are specifically defective in the calcium-dependent deactivation mechanism. These data suggest that the light-dependent calcium influx inactivates this cascade through a biochemical pathway that requires the *inaC* gene product, and that this mechanism represents a molecular basis for stimulus-dependent regulation of visual transduction in *Drosophila* photoreceptors.

The compound eye of *Drosophila* is composed of 800 ommatidia, or unit eyes, each of which contains eight photoreceptor neurons⁶. These eight cells can be divided into three classes: R1–R6, R7 and R8 (ref. 7). R1–R6 cells express a blue-absorbing rhodopsin, Rh1 (refs 8, 9), and are the major photoreceptor cell class in the retina. We have focused on the analysis of R1–R6 cells, as they are the best characterized class of photoreceptors with respect to signal transduction, and are the most experimentally accessible cell type. *Drosophila* photoreceptors were dissociated from the heads of late pupae 1–5 h before eclosion (stages p14, p15)¹⁰ in clusters of R1–R8 cells, each of which represents one ommatidium stripped of all support and pigment cells (Fig. 1). A similar preparation from *Drosophila* adults has been described by Hardie *et al.*¹¹. The R1–R6 cell soma are easily identified by the unique location of their nuclei at the very distal margin of each cluster (Fig. 1c).

Short flashes of white light (10-ms duration) used to stimulate dark-adapted photoreceptors evoke large currents with an average latency of 24.9 ± 2.3 ms ($n = 16$) (Fig. 1d), which under strong stimulation can reach more than 20 nA at peak amplitude. Table 1 summarizes a series of ion-substitution experiments demonstrating that the *Drosophila* light-activated channel is a cation-selective channel that can pass even large monovalent ions such as TEA⁺ and Tris⁺, suggesting a fairly large pore size. In addition, raising the extracellular Ca²⁺ concentration ([Ca²⁺]_{out}) to 20 from 0.1 mM resulted in a +28 mV change in the reversal potential, indicating that Ca²⁺ passes through the pore with a high relative permeability.

It is significant that Ca²⁺ permeates this light-activated conductance because Ca²⁺ can act as a messenger mediating adaptation in vertebrate phototransduction^{12–14} and may also be important in both excitation^{15,16} and regulation of invertebrate

TABLE 1 Summary of ion substitution experiments

Bath [Ca ²⁺] (mM)	Bath monovalents	Pipette monovalents	E_{rev} (mV)
0.1	124 mM CsCl	124 mM CsCl	-0.82 ± 0.75
0.1	124 mM NaCl	124 mM CsCl	-3.33 ± 0.4
0.1	124 mM KCl	124 mM CsCl	-2.18 ± 1.03
0.1	124 mM Tris-Cl	124 mM CsCl	-9.79 ± 0.43
0.1	124 mM TEA-Cl	124 mM CsCl	-24.76 ± 2.77
0.1	90 mM Cs-gluconate, 34 mM CsCl	124 mM CsCl	-0.981 ± 0.606
20	124 mM CsCl	124 mM CsCl	$+28.74 \pm 1.43$

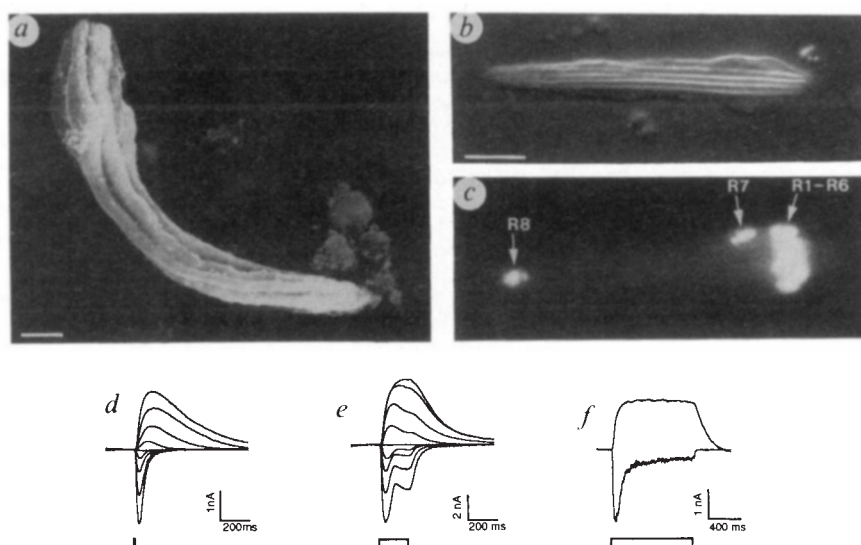
Pipette solution contained 124 mM CsCl, 10 mM HEPES, 11 mM EGTA, 1 mM CaCl₂, 2 mM MgCl₂, 3 mM Mg²⁺-ATP, 0.5 mM Na⁺-GTP, pH 7.15 with CsOH. Bath solutions contained 124 mM monovalents as indicated, 10 mM HEPES, CaCl₂ as indicated, 32 mM sucrose, pH 7.15. Each value of E_{rev} is the average $\pm$ standard deviation of data from at least three cells. Reversal potentials were calculated from plotting peak currents against holding potential. The sequence of ion permeabilities was determined from reversal potentials using the Goldman-Hodgkin-Katz^{28,29} equation simplified for the case of two permeant ions at the same intracellular(i) and extracellular(o) concentration: $E_{rev} = RT/zF \ln(P_o/P_i)$. Relative permeabilities of ions: $P_{Ca^{2+}} > P_{Cs^{+}} > P_{Na^{+}} \sim P_{K^{+}} > P_{Tris^{+}} > P_{TEA^{+}}$.

phototransduction^{17–19}. Pulses of light (200 ms or 1 s duration; Fig. 1e, f) evoke complex responses that have biphasic kinetics resulting from a rapid partial desensitization (attenuation of current) during the stimulus. The responses both to flashes and to pulses of light show a marked asymmetry in the kinetics of the light-activated currents about the reversal potential. For example, the kinetics of the rapid desensitization (as defined above) and deactivation (recovery of the light-activated response after termination of the stimulus) are much faster during inward current flow (Fig. 1d, e). The asymmetries in the kinetics of the light-activated currents seem not to be directly voltage dependent, but instead to be dependent on the direction of ion flow. Thus shifting the reversal potential under standard extracellular calcium concentrations (1 mM) simply causes the same asymmetries to be reflected around the new reversal potential (data not shown). Similar results obtained with the use of the nystatin perforated-patch technique²⁰ (Fig. 1f) demonstrate that these asymmetries are not the result of nonphysiological alterations of the intracellular environment during whole-cell recording. These results suggest a model in which extracellular calcium entering through the light-activated conductance may regulate the signalling cascade and may be responsible for the rapid kinetics of deactivation and desensitization of the response.

As this model predicts that the kinetics of the light response should be a function of [Ca²⁺]_{out}, we examined the effects of different concentrations of extracellular calcium on the rates of the activation (defined as rate of onset of current after stimulus) and deactivation processes. Varying [Ca²⁺]_{out} from 1 mM to 20 mM causes a large increase in the rate of both activation and deactivation of the inward currents, but has a much smaller effect on the outward currents (Fig. 2b, c). In addition, both the rate and the degree of the rapid desensitization during a sustained light stimulus are greatly increased when the photoreceptors are exposed to 20 mM [Ca²⁺]_{out} (compare Fig. 3b with c). These changes are expected to affect predominantly the inward currents if only inward currents carry large amounts of Ca²⁺ into the cell, and this calcium entry then regulates the activation, deactivation and desensitization processes. During the outward currents, the calcium influx into the cell is presumably greatly attenuated both by the smaller electrochemical driving force and the opposing bulk flow of monovalent ions. For quantitative evaluation of the inward current data, the rate of activation was measured as the time to 50% of peak current from onset of the light stimulus, and the rate of the deactivation was measured

¶ To whom correspondence should be addressed.

FIG. 1 Isolated *Drosophila* photoreceptor clusters, each representing one ommatidium, and whole-cell voltage-clamp recordings of light-activated currents. **a**, Scanning electron micrograph (SEM) showing that each cluster comprises a group of tightly packed individual photoreceptors. Scale bar, 10 μm . **b**, Normarski interference contrast image of a photoreceptor cluster; the distal end of the photoreceptors is to the right. Scale bar, 25 μm . **c**, The same cluster as in **b**, stained with a fluorescent marker specific for nuclei. The arrows indicate the position of the R1–R6, R7 and R8 nuclei. **d**, Light-activated currents evoked by 10 ms flashes of white light at holding potentials from -80 mV to $+80\text{ mV}$. Leak subtractions were done off-line, and all traces were normalized to zero at the pretrigger region just before onset of the light stimulus. Recordings of light-activated currents were made in symmetric Cs^+ solutions (to block voltage-activated K^+ -channels, and aid in the isolation of the light-activated currents) with 1 mM Ca^{2+} in the bath. **e**, Light-activated currents evoked by 200-ms pulses of white light with the same protocol as in **d**. **f**, Light-activated currents evoked by 1-s pulses of white light at -80 mV and $+80\text{ mV}$ holding potentials. Records in **f** were obtained with the use of the nystatin perforated patch technique. All light-activated responses were obtained from different cells (in all figures).



METHODS. Heads were dissected from stage p15 wild-type Canton S *Drosophila* pupae, and rapidly chopped in 5 μl Ca^{2+} -free, Mg^{2+} -free Ringer's solution (120 mM NaCl, 4 mM KCl, 10 mM HEPES, 32 mM sucrose, pH 7.15) on a glass plate. The resulting suspension was then transferred to about 30 μl of the same solution, and gently triturated through a pipette tip. All manipulations were done under dim red lights to prevent photoreceptor desensitization. SEM was done using standard techniques. For fluorescent labelling of nuclei, the mixture was fixed in 2% glutaraldehyde for 10 min, incubated with 0.1% saponin (Sigma) for 1 min and incubated with 100 ng ml^{-1} Hoechst 33258 nuclear stain. Isolated photoreceptor clusters were immediately photographed without further manipulation. All processing steps were done at room temperature (23–25 $^{\circ}\text{C}$). For electrophysiology, dissociated photoreceptor clusters were immediately allowed to settle under the bath solution onto a clean glass coverslip forming the bottom of a 35-mm-diameter recording dish. The dish was mounted onto the stage of a Leitz Fluovert inverted microscope, and cells were visualized through Hoffmann interference contrast under weak red-light illumination. The isolated photoreceptor clusters are stable in culture for several hours. In the whole-cell recording configuration, most late pupal photoreceptors had input

resistances of 200–1,500 $\text{M}\Omega$, and whole-cell capacitances of 35–50 pF. Recordings (23–25 $^{\circ}\text{C}$) were made with patch pipettes made from fibre-filled borosilicate glass (Sutter Instruments) with average resistances of 5 $\text{M}\Omega$. Junction potentials were nulled just before seal formation, and no changes in the ion composition of the bath were made after seal formation. Whole-cell recordings were made using standard techniques²⁸. Perforated-patch recording techniques were adapted from ref. 20. Recordings were made with a 50 mg ml^{-1} stock nystatin (Sigma) solution in DMSO freshly diluted 1:500 into filtered pipette solution for each pipette. Pipette tips were filled with nystatin-free solution to aid in seal formation. Most (80%) series-resistance errors were compensated during recording, and data was only acquired from cells with seal resistances $>5\text{ G}\Omega$. Signals were amplified with an Axopatch 1-D patch-clamp amplifier (Axon Instruments), filtered through an 8-pole Bessel filter at 2 kHz, and digitized at 125 kHz for analysis. Data were analysed using pClamp 5.5.1 software (Axon Instruments). Photoreceptors were stimulated with light from a 75 W Xenon arc lamp, attenuated using neutral density filters (Oriel), regulated by an electronic shutter and focused through the microscope objective. Physiological electrode solution: 120 mM KCl, 4 mM NaCl, 10 mM HEPES, 11 mM EGTA, 1 mM CaCl_2 , 2 mM MgCl_2 , 3 mM Mg^{2+} -ATP, 0.5 mM Na^+ -GTP, pH 7.15 with KOH. Physiological bath solution: 120 mM NaCl, 4 mM KCl, 10 mM HEPES, 1 mM CaCl_2 , 32 mM sucrose, pH 7.15 with NaOH. Symmetric Cs^+ solutions were as above, except that pipette and bath monovalent cations were replaced with 124 mM CsCl.

as the time from peak response to 75% attenuation of the peak current. Figure 2g and h shows these measurements plotted against $[\text{Ca}^{2+}]_{\text{out}}$ (solid bars). The results indicate that it is possible to manipulate the kinetics of activation, deactivation and desensitization of the light-activated response as a function of $[\text{Ca}^{2+}]_{\text{out}}$, thus demonstrating that extracellular calcium is sufficient to regulate the light response.

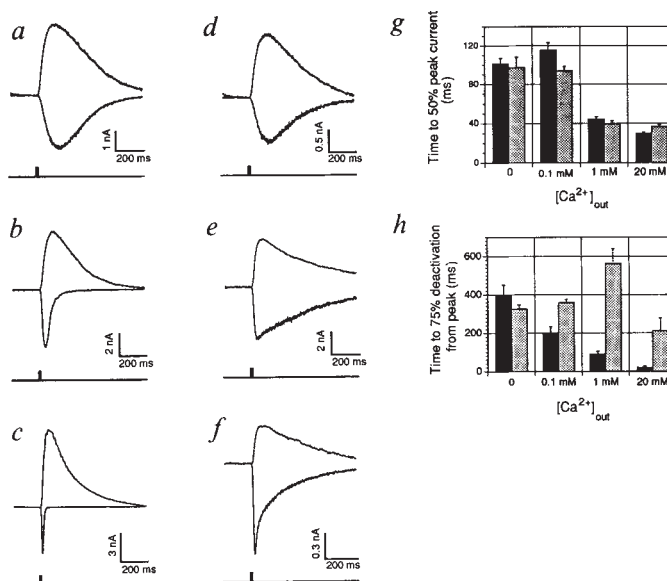
To demonstrate the necessity of extracellular calcium in these processes, we recorded from cells under conditions in which $[\text{Ca}^{2+}]_{\text{out}}$ is buffered with EGTA to $<10^{-8}\text{ M}$. In concordance with our model, these cells not only have extremely slow activation and deactivation rates, but now the kinetics of the currents become symmetric and independent of direction of ion flow (Fig. 2a). Additionally, the kinetics of these currents no longer show any biphasic characteristics in response to a pulse of light (Fig. 3a), suggesting that the rapid desensitization also requires extracellular calcium. Extracellular Ca^{2+} must enter the photoreceptor to regulate the light response, because buffering intracellular calcium levels with a fast calcium chelator, BAPTA (1,2-bis(*O*-aminophenoxy)ethane-*N,N,N',N'*-tetraacetic acid) (ref. 21), instead of the slower EGTA, specifically reduces the rates of activation and deactivation (compare Fig. 3c with f). But the degree of response desensitization is similar to either calcium buffer, consistent with the fact that EGTA and BAPTA have similar dissociation constants for Ca^{2+} . Taken together,

these results show that calcium influx through the light-activated conductance is required for regulating the light response.

To investigate the molecular bases of these calcium-dependent mechanisms, we screened known phototransduction mutants for those that may be deficient in these processes. Many of these mutants were isolated through mutagenic screens for flies with defective electroretinograms (ERG)^{5,22}. ERGs are extracellular recordings of the electrical activity of the whole eye in response to light²³. Figure 2d, e and f show the light-activated responses of homozygous *inaC* photoreceptors^{5,24} at various $[\text{Ca}^{2+}]_{\text{out}}$. These photoreceptors are unable to carry out normal rapid deactivation after a light stimulus. By contrast, excitation mechanisms seem to be unaffected in this mutant, as the dependence of the rate of activation on $[\text{Ca}^{2+}]_{\text{out}}$ in *inaC* photoreceptors is similar to wild type (Fig. 2g, compare solid and dotted bars). To exclude the possibility that the *inaC* phenotype is caused by the inability of calcium to enter the photoreceptor cells, we showed that, as in wild-type photoreceptors, the reversal potential of the light-activated conductance in the mutant shifts with increasing $[\text{Ca}^{2+}]_{\text{out}}$ (not shown). The outward currents are largely unaffected by the mutation, suggesting that the expression of the mutant phenotype is dependent on the direction of ion flow. Figure 2h (dotted bars) shows the quantitative analysis of the inward current deactivation kinetics in *inaC* photoreceptors, demonstrating that *inaC* mutants are

FIG. 2 Kinetics of activation and deactivation of the light-activated currents in wild-type and *inaC* mutant photoreceptors with varying $[Ca^{2+}]_{out}$. All photoreceptors were recorded under conditions of symmetric Cs^+ solutions (see Fig. 2 legend) with varying concentrations of extracellular calcium. *a, b, c*, Responses of wild-type cells with 0, 1 mM and 20 mM $[Ca^{2+}]_{out}$, respectively; *d, e, f*, responses of *inaC* mutant photoreceptors with 0, 1 mM and 20 mM $[Ca^{2+}]_{out}$, respectively. *a–f*, Recordings taken at holding potentials of -80 mV and $+80$ mV. *g*, Plot of time to 50% of peak current from onset of light stimulus for inward currents against $[Ca^{2+}]_{out}$. *h*, Plot of time from peak to 75% attenuation of the peak amplitude for inward currents against $[Ca^{2+}]_{out}$. Both graphs show data from wild-type (solid bars), and *inaC* mutant (dotted bars) photoreceptors. *g, h*, Recordings were done at a holding potential of -80 mV, and the stimulus consisted of a 10-ms flash of white light. Each data point consists of averaged data from at least five cells, and data from each cell consists of an average of at least nine individual traces.

METHODS. All wild-type data shown were acquired from Canton S photoreceptors. All mutant data shown were acquired from photoreceptors homozygous for either *inaC*²⁰⁷ or *inaC*²⁰⁹ alleles. All experiments were repeated with both alleles, and no phenotypic differences were observed between them (data not shown).



specifically defective in the calcium-dependent deactivation mechanism. Interestingly, *inaC* mutants are also defective in the rapid desensitization to longer light pulses (compare Fig. 3*b* and *e*). These findings suggest that Ca^{2+} influx as a result of channel activation causes rapid deactivation and desensitization through a regulatory pathway that depends on the *inaC* gene product.

To test for the requirement of extracellular calcium in regulating *inaC* activity, we compared the light response of mutant and wild-type photoreceptors under conditions in which

$[Ca^{2+}]_{out}$ is buffered to $<10^{-8}$ M (Fig. 2*a, d, h*). Under these conditions, wild-type photoreceptors have deactivation kinetics similar to those of *inaC* cells. As the expression of the mutant phenotype requires the presence of extracellular Ca^{2+} , we conclude that wild-type *inaC* activity is dependent on the entry of extracellular calcium through the light-activated conductance.

Calcium-mediated mechanisms underlie many aspects of regulation of the visual cascade in vertebrate photoreceptors¹. Recently, Ca^{2+} -binding proteins that seem to regulate guanylate cyclase^{25,26} and cyclic GMP phosphodiesterase²⁷ have been purified and biochemically characterized. The ability to study *Drosophila* mutants with specific defects in calcium-dependent mechanisms involved in the regulation of the phototransduction cascade provides a powerful approach for the molecular investigation of these processes. □

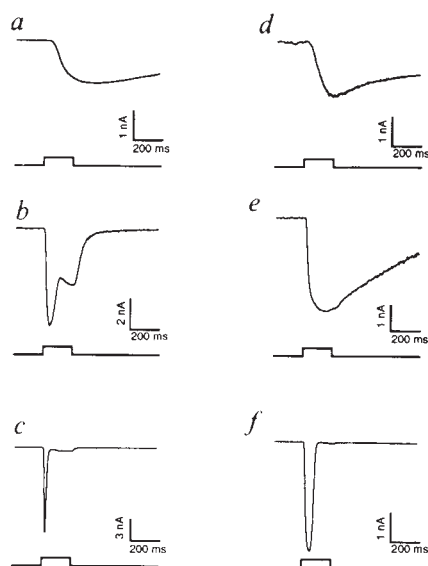


FIG. 3 Kinetics of desensitization during prolonged light stimuli in wild-type and *inaC* mutant photoreceptors with varying $[Ca^{2+}]_{out}$. *a, b, c*, Responses of wild-type photoreceptors recorded in symmetric Cs^+ solutions at a holding potential of -80 mV under conditions of 0, 1 mM and 20 mM $[Ca^{2+}]_{out}$, respectively. *d, e*, Responses of *inaC* photoreceptors recorded under the same conditions with 0 and 1 mM $[Ca^{2+}]_{out}$, respectively. The stimulus was a 200-ms pulse of white light. *f*, Response of a wild-type photoreceptor recorded under identical conditions as in *c* (20 mM $[Ca^{2+}]_{out}$, symmetric Cs^+ solutions) at a holding potential of -80 mV, except that buffering of Ca^{2+} levels in the pipette was with BAPTA (Molecular Probes) instead of EGTA. Pipette solution was as in Fig. 1 legend, except that EGTA was replaced with the same concentration of BAPTA.

Received 13 May; accepted 15 October 1991.

1. Fain, G. L. & Matthews, H. R. *Trends Neurosci.* **13**, 378–384 (1990).
2. Pugh, E. N. & Lamb, T. D. *Vision Res.* **30**, 1923–1948 (1990).
3. Payne, R. *Photobiochem. Photobiophys.* **13**, 373–397 (1986).
4. Fein, A. & Payne, R. in *Facets of Vision* (eds Stavenga, D. & Hardie, R.) 173–185 (Springer, Berlin, 1989).
5. Pak, W. L. in *Neurogenetics, Genetic Approaches to the Nervous System* (ed. Breakfield, X. O.) 67–99 (Elsevier, New York, 1979).
6. Tomlinson, A. *Development* **104**, 183–193 (1988).
7. Hardie, R. C. in *Progress in Sensory Physiology* (ed. Ottoson, D.) 1–79 (Springer, New York, 1983).
8. O'Tousa, J. E. et al. *Cell* **40**, 839–850 (1985).
9. Zuker, C. S., Cowman, A. F. & Rubin, G. M. *Cell* **40**, 851–858 (1985).
10. Bainbridge, S. P. & Bownes, M. *J. Embryol. exp. Morph.* **66**, 57–80 (1981).
11. Hardie, R. C., Voss, D., Pongs, O. & Laughlin, S. B. *Neuron* **6**, 477–486 (1991).
12. Matthews, H. R., Murphy, R. L. W., Fain, G. L. & Lamb, T. D. *Nature* **334**, 67–69 (1988).
13. Nakatani, K. & Yau, K.-W. *Nature* **334**, 69–71 (1988).
14. Koch, K. & Stryer, L. *Nature* **334**, 64–66 (1988).
15. Rubin, L. J. & Brown, J. E. *Biophys. J.* **47**, 38 (1985).
16. Payne, R., Corson, D. W., Fein, A. & Berridge, M. J. *J. gen. Physiol.* **88**, 127–142 (1986).
17. Brown, J. E. & Blinks, J. R. *J. gen. Physiol.* **64**, 643–665 (1974).
18. Fein, A. & Lisman, J. E. *Science* **187**, 1094–1096 (1975).
19. Lisman, J. E. & Brown, J. E. *J. gen. Physiol.* **59**, 701–719 (1972).
20. Horn, R. & Marty, A. *J. gen. Physiol.* **92**, 145–159 (1988).
21. Tsien, R. Y. *Biochemistry* **19**, 2396–2404 (1980).
22. Koenig, J. & Merriam, J. *Drosoph. Inf. Serv.* **52**, 50–51 (1977).
23. Hengstenberg, R. & Gotz, K. G. *Kybernetik* **3**, 276 (1967).
24. Pak, W. L. in *Handbook of Genetics* (ed. King, R. C.) 703–733 (Plenum, New York, London, 1975).
25. Dizhoor, A. M. et al. *Science* **251**, 915–918 (1991).
26. Lambrecht, H. G. & Koch, K. W. *EMBO J.* **10**, 793 (1991).
27. Kawamura, S. & Murakami, M. *Nature* **349**, 420–423 (1991).
28. Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. *Pflügers Arch.* **391**, 85–100 (1981).

ACKNOWLEDGEMENTS. We thank W. Pak, J. Merriam and R. Hardy for *inaC* alleles; R. Y. Tsien, W. Harris, L. Stryer, D. Baylor and members of the Zuker lab for advice on the manuscript; M. P. Mahaut-Smith for help with the perforated patch technique; and P. O'Donnell for assistance with electron microscopy. This work was supported by the Howard Hughes Medical Institute. C.S.Z. acknowledges support from the Pew Foundation, the McKnight Foundation, the Alfred P. Sloan Foundation and the Basil O'Connor program from the March of Dimes. R.R. was supported by a training grant from the Medical Scientist Training Program. G.L.H. acknowledges support from the National Eye Institute and the NSF. C.F.S. and C.S.Z. are investigators of the Howard Hughes Medical Institute.

CD43, a molecule defective in Wiskott–Aldrich syndrome, binds ICAM-1

Yvonne Rosenstein^{*†}, John K. Park^{*}, William C. Hahn^{*}, Fred S. Rosen^{‡§}, Barbara E. Bierer^{*||} & Steven J. Burakoff^{*‡}

^{*} Division of Pediatric Oncology, Dana-Farber Cancer Institute, [§] The Center for Blood Research, ^{||} Hematology–Oncology Division, Brigham and Women's Hospital, and the Departments of [†] Pathology, [‡] Pediatrics, and ^{||} Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA

THE protein CD43 (also known as sialophorin, leukosialin, large sialoglycoprotein or gp115) is expressed on the surface of T lymphocytes, monocytes, neutrophils, platelets and some B lymphocytes^{1–6}. Expression of CD43 is deficient and/or defective in the X-chromosome-linked immunodeficiency disorder Wiskott–Aldrich syndrome⁷, suggesting that CD43 might have a role in T-cell activation. We have shown that expression of human CD43 in an HLA-DR-specific murine T-cell hybridoma enhances the antigen-specific response to stimulation by the human lymphoblastoid cell line Daudi, and that Daudi cells bind specifically to purified immobilized CD43 (ref. 8). These data indicate that the specific interaction of CD43 with a ligand on the surface of Daudi cells might contribute to T-cell activation. Here we report evidence that intercellular adhesion molecule-1 (ICAM-1, or CD54), is a ligand for CD43.

To identify a ligand for CD43, we screened a series of monoclonal antibodies raised against B-lymphocyte surface antigens (from the IVth International Workshop on human leukocyte differentiation antigens) for their ability to inhibit the binding of Daudi cells to plate-bound purified CD43. Only three out of 56 antibodies consistently inhibited the interaction of Daudi cells with purified CD43: LB.2, RR.1 and F2B1, and all three of these recognize the ICAM-1 molecule^{9–11}. As shown in Fig. 1, binding of Daudi cells to CD43 was inhibited ~50% by the anti-CD43 antibody G10-2 (ref. 12) (or B1B6; ref. 13, and data not shown), as well as with the anti-ICAM-1 antibody LB.2. However, other anti-CD43 monoclonal antibodies, G19-1 (ref. 12) and L10 (ref. 4) were unable to inhibit binding. Binding was not affected by the irrelevant antibody, anti-CD2 TS2/18 (ref. 14). These data indicated that ICAM-1 could be a ligand for CD43.

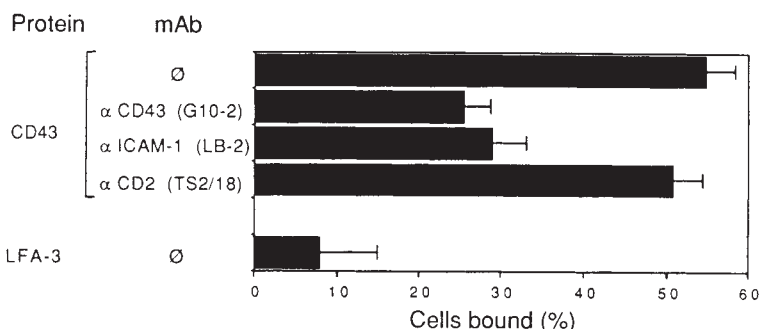
The reciprocal experiment was performed to assess the binding of CD43⁺ cells to immobilized recombinant human ICAM-1. We analysed murine T-cell hybridomas expressing either human CD43 (ref. 8) or, as a control, human CD2 (ref. 15). CD43⁺ cells and CD2⁺ cells were tested for their ability to bind to immobilized recombinant ICAM-1 (ref. 16) or immobilized immunoaffinity-purified LFA-3 (ref. 17). CD43⁺ hybridomas bound to ICAM-1 but not to LFA-3 (Fig. 2a); CD2⁺ cells bound to LFA-3 but not to ICAM-1 (Fig. 2b). Anti-ICAM-1 (F2B1.8, LB.2, RR1/1) antibodies and anti-CD43 (G10-2) antibody inhibited binding of CD43⁺ hybridomas to ICAM-1 and had no effect on the binding of CD2⁺ hybridomas to LFA-3 (Fig. 2b, and data not shown). Confirming the specificity of this interaction, neither anti-CD2 (TS2/18) nor anti-LFA-3 (TS2/9; ref. 14) antibodies inhibited the CD43/ICAM-1 interaction; both, however, abolished the adhesion of the CD2⁺ hybridomas to immobilized LFA-3. Although murine LFA-1 is reported not to bind human ICAM-1 (ref. 18), all T-cell hybridomas were preincubated with the anti-LFA-1 β -chain antibody M-17/4.2 (ref. 19) to ensure that murine LFA-1 did not participate in binding.

To eliminate the possibility that other cell-surface molecules might be involved in the CD43/ICAM-1 interaction, we incorporated these molecules onto separate cell-sized beads (tosyl-activated magnetic beads and silica beads) to assess conjugate formation. Immunoaffinity-purified CD43 was bound to the surface of the silica beads and recombinant ICAM-1 or immunoaffinity purified HLA-DR molecules were coupled to the tosyl-activated magnetic beads. The expression of CD43, HLA-DR, and ICAM-1 was assessed by indirect immunofluorescence (Fig. 3a) and the ability of the CD43⁺ beads to form conjugates with beads expressing either ICAM-1 or HLA-DR was evaluated (Fig. 3b). Incubation of the CD43⁺ beads with ICAM-1⁺ beads resulted in the formation of 42% conjugates, whereas conjugate formation between CD43⁺ beads and HLA-DR⁺ beads was ~10%. CD43/ICAM-1 conjugate formation was inhibited by the anti-CD43 antibody G10-2 (60% inhibition) or anti-ICAM-1 antibody LB.2 (70% inhibition), but not by the irrelevant antibodies anti-CD2 (TS2/18) and anti-LFA-3 (TS2/9). The ability of the CD43⁺ beads to form conjugates with the ICAM-1⁺ beads provides direct evidence that CD43 and ICAM-1 constitute a receptor–ligand pair, thus defining a novel pathway of cell–cell adhesion.

ICAM-1 is a ligand for the integrins LFA-1 (CD11a/CD18; ref. 20) and CR3 (CD11b/CD18; ref. 21). These integrins are heterodimeric proteins; the former is involved in lymphoid-cell

FIG. 1 Binding of Daudi cells to immobilized CD43 is inhibited by anti-ICAM-1 monoclonal antibody (mAb). The binding to immobilized CD43 of Daudi, a human Burkitt's lymphoma-derived B-cell line, was assessed in the presence of different mAbs: G10-2 (anti-CD43, IgG1), LB.2 (anti-ICAM-1, IgG2a), and TS2/18 (anti-CD2, IgG1). Results are expressed as a percentage of specific binding; each value represents the mean of 6 replicates. The experiment shown is representative of three independent experiments. Φ , Wells coated with purified protein, with no mAb.

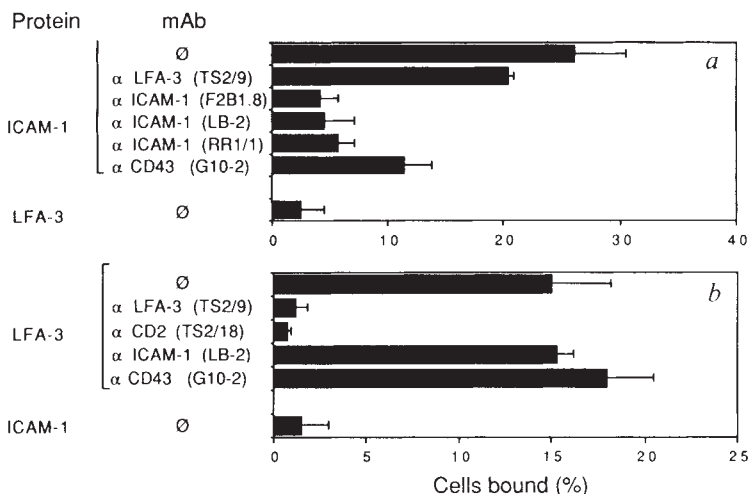
METHODS. CD43 was immunoaffinity-purified and immobilized on 96-well plates (Nunc, Norway) from ¹²⁵I-labelled and unlabelled JY-cell lysate using an L10 (anti-CD43) mAb-conjugated Sepharose-4B column as described⁸. Equal amounts of immunopurified LFA-3, isolated as before¹⁷, were added to control wells under identical conditions. Plates were incubated for 3 h at 37 °C, after which the remaining protein-binding sites were blocked with 1% BSA for 2 h at 37 °C. Wells were extensively washed to remove unadsorbed proteins. For the mAb inhibition assays, where indicated, protein-coated wells were incubated with G10-2 mAb at 1 μ g ml⁻¹ for 4–6 h at 37 °C; excess antibody was removed before the addition of cells. [³H]Thymidine-labelled Daudi cells, preincubated for 15 min with anti-CD18 mAb (TS1/18) to eliminate homotypic adhesions, were treated with either LB.2 or TS2/18 mAbs at 1 μ g ml⁻¹ for an additional 30 min. Cells were washed with RPMI 1640 medium containing 10% bovine calf serum (BCS) before



adding them to the protein-coated plates (2×10^5 per well). Plates were centrifuged at 50g for 1 min and incubated at 37 °C for 30–40 min. Non-adherent cells were removed by aspirating the fluid contents of each well through an 18-gauge needle and carefully rinsing each well twice with 200 μ l of prewarmed (37 °C) RPMI 1640 10% with BCS. After washing, the radioactivity in each well was counted. The cells bound were expressed as a percentage and calculated as: [(c.p.m. of cells retained by CD43- or LFA-3-adsorbed wells) – (c.p.m. of cells retained by BSA-adsorbed wells)] $\div$ [(total c.p.m. of cells added to the wells) – (c.p.m. of cells retained by BSA-adsorbed wells)].

FIG. 2 CD43⁺ cells bind immobilized recombinant soluble ICAM-1. *a*, Binding of ³H-labelled CD43⁺ (16CD43-3.82), or *b*, CD2⁺ (16CD2-XH14) hybridoma cells to purified ICAM-1 or LFA-3 was measured by incubating cells on ICAM-1 or LFA-3-coated microtitre wells for 30 min at 37 °C. Results are the mean of triplicate samples. The experiment shown is representative of four independent experiments.

METHODS. Fifty microlitres of a solution with 2 µg ml⁻¹ of recombinant soluble ICAM-1 in Dulbecco's PBS was added to the wells of a microtitre plate (Linbro-Titertek, Flow Laboratories) for 1 h. Control wells were coated with affinity-purified human LFA-3 (25 µg per well)¹⁷ for the same length of time. After blocking the remaining protein binding sites with 2% BSA in Dulbecco's PBS (200 µl per well) for 2 h at 37 °C, plates were washed five times with Dulbecco's PBS and once with RPMI 1640 with 10% BCS. To the appropriate wells, mAbs (anti-ICAM-1 F2B1, LB.2 and RR1/1, or anti-LFA-3 TS2/9, at 1 µg ml⁻¹) were added overnight, and excess mAb was washed off before the addition of cells. Hybridoma cells were pretreated with anti-murine CD11a mAb (M17/4.2) for 15 min and for an additional 30 min with either G10-2 (anti-CD43) or TS2/18 (anti-CD2) mAbs at 1 µg ml⁻¹. Cells were washed to remove excess antibody and plated at 2 × 10⁵ cells per well. After 30 min incubation at 37 °C, nonadherent cells were removed by repetitive pipetting and the radioactivity left in each well was measured. The



percentage of cells bound was calculated as described in the legend to Fig. 1.

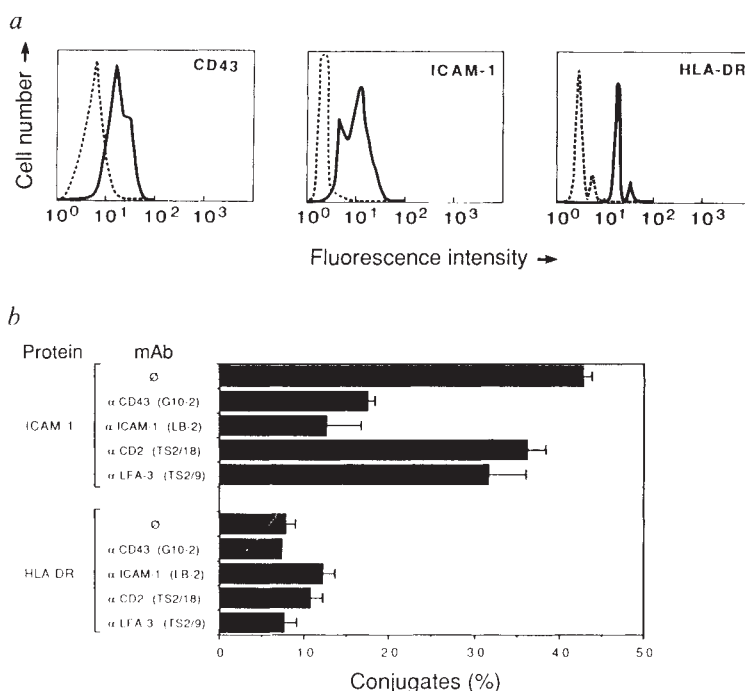
adhesion and the latter is the receptor for C3b and plays a part in myeloid cell adhesion²². The binding site on ICAM-1 for LFA-1 (and rhinovirus) maps to the first amino-terminal immunoglobulin-like domain²³ and differs from that for CR3, which maps to the third N-terminal immunoglobulin-like domain²⁴. The data presented here are evidence that ICAM-1, a member of the immunoglobulin superfamily of adhesion receptors, is a ligand for CD43, a single-chain molecule with no structural similarity to integrin family members. Expressed on all T cells, a subset of B cells and other haematopoietic cells, CD43 may be important for cellular activation and lymphoid cell function, because anti-CD43 antibodies have a stimulatory or co-stimulatory effect on these cells *in vitro*^{9,13,25-28}. Thus, the interaction of CD43 with ICAM-1 may result not only in cell-cell adhesion but also in T cell, natural killer cell, B cell and monocyte signal transduction through CD43, a function that

has been more difficult to verify for the integrins LFA-1 and CR3. As a result of its extended conformation²⁹ allowing it to protrude from the cell surface, CD43 is likely to be important in lymphoid cell interactions with ICAM-1.

Patients suffering from Wiscott-Aldrich syndrome have a profound immunodeficiency in T-cell function. As these patients either do not express CD43 or express different forms of it, their immunodeficiency may result from defective T-cell adhesion and/or activation mediated through the CD43/ICAM-1 interaction. The observation that a significant percentage of individuals with human immunodeficiency virus infection have circulating anti-CD43 antibodies has led to the suggestion that these autoantibodies may contribute to the severe immunodeficiency found in AIDS patients³⁰. These clinical data compell us to investigate further the role of CD43/ICAM-1 interactions in lymphoid cell function. □

FIG. 3 Purified CD43 directly interacts with recombinant ICAM-1. *a*, Surface expression of membrane proteins on the surface of the silica beads and the magnetic beads was assessed by indirect immunofluorescence. Beads were stained with either anti-CD43 (L10), anti-ICAM-1 (LB.2), or anti-HLA-DR (LB3.1) (ref. 31) mAb before the addition of a fluorescein-conjugated rabbit anti-mouse IgG antibody. *b*, Conjugate formation between CD43-coated beads and ICAM-1- or HLA-DR-coated beads was assessed after 1 h incubation at 37 °C. Data shown are representative of three independent experiments.

METHODS. Spherisorb 5-µm ODS1 silica beads (Phase Sep) were reconstituted with protein as described³², except that lipids were added as a mixture³³. Dynabeads M-450 tosyl-activated were the support for ICAM-1 or HLA-DR. These polystyrene paramagnetic beads are, like the silica beads, uniform in size (4.5 µm) and they are activated by treatment with *p*-toluenesulphonyl chloride for chemical binding of proteins. Proteins were bound to the beads according to the instructions of the manufacturer. After removal of unbound protein, all beads were stored at 4 °C in Dulbecco's PBS, 0.1% BSA with sodium azide, and washed twice before use. For adhesion assays, CD43⁺ beads were incubated with ICAM-1⁺ beads or HLA-DR beads at a 4:1 ratio in a final volume of 30 µl RPMI 1640 with 10% BCS. To the appropriate samples, mAbs (at 1 µg ml⁻¹) were added at the start of the experiment. After incubation at 37 °C, samples were gently resuspended 4 to 6 times with a pipette tip and conjugates were counted under a light microscope. A conjugate was defined as the binding of one dynabead to at least one silica bead. Dynabeads containing ICAM-1 or HLA-DR were identified by their brown colour (silica beads are white). Samples were analysed in duplicates, and two or more



counts of at least 100 M-450 dynabeads were performed for each sample and scored as conjugates or non-conjugates. Results are expressed as mean value ± s.e.m. of four counts made for each group.

Received 12 August; accepted 17 September 1991.

1. Gamberg, C. G., Hayry, P. & Anderson, L. C. *J. Cell Biol.* **68**, 642-653 (1976).
2. Brown, W. R. A., Barclay, N., Sunderland, C. A. & Williams, A. *Nature* **289**, 456-460 (1981).
3. Carlsson, S. R. & Fukuda, M. *J. Biol. Chem.* **261**, 12779-12786 (1986).
4. Remold-O'Donnell, E., Zimmerman, C., Kenney, D. & Rosen, F. *Blood* **70**, 104-109 (1987).
5. Borche, L., Lozano, F., Vilella, R. & Vives, J. *Eur. J. Immunol.* **17**, 1523-1526 (1987).
6. Wicken, M., Bjorck, P., Axelsson, B. & Perlmann, P. *Scand. J. Immunol.* **28**, 457-464 (1988).
7. Parkman, R., Remold-O'Donnell, E., Kenney, D., Perrine, S. & Rosen, F. *Lancet* **ii**, 1387-1389 (1981).
8. Park, J. K. *et al. Nature* **350**, 706-709 (1991).
9. Clark, E. A., Ledbetter, J. A., Holly, R. C., Dimmorf, P. A. & Shu, G. *Hum. Immunol.* **16**, 100-113 (1986).
10. Rothlein, R., Dustin, M. L., Marlin, S. D. & Springer, T. A. *J. Immunol.* **137**, 1270-1274 (1986).
11. Dorken, B., Moller, P., Pezzutto, A., Schwartz-Albiez, R. & Moldenhauer, G. in *Leucocyte Typing IV* (ed. Knapp, W.) 3-32 (Oxford University Press, 1989).
12. Stoll, M., Dalchau, R. & Schmidt, R. in *Leucocyte Typing IV* (ed. Knapp, W.) 605-608 (Oxford University Press, 1989).
13. Axelsson, B., Youseffi-Etemad, R., Hammarstrom, S. & Perlmann, P. *J. Immunol.* **141**, 2912-2917 (1988).
14. Sanchez-Madrid, F. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 7489-7493 (1982).
15. Bierer, B. E., Peterson, A., Barbosa, J., Seed, B. & Burakoff, S. J. *Proc. natn. Acad. Sci. U.S.A.* **85**, 1154-1158 (1988).
16. Marlin, S. O. *et al. Nature* **344**, 70-72 (1990).
17. Hahn, W. C., Rosenstein, Y., Burakoff, S. J. & Bierer, B. E. *J. Immunol.* **147**, 14-21 (1991).
18. Johnston, C., Dustin, M., Hibbs, M. & Springer, T. A. *J. Immunol.* **145**, 1181-1187 (1990).

19. Sanchez-Madrid, F., Davignon, D., Martz, E. & Springer, T. A. *Cell Immunol.* **73**, 1-11 (1982).
20. Springer, T. A., Dustin, M. L., Kishimoto, T. K. & Marlin, S. D. *A. Rev. Immunol.* **223-252** (1987).
21. Diamond, M. S. *et al. J. Cell Biol.* **111**, 3129-3139 (1990).
22. Larson, R. S. & Springer, T. A. *Immunol. Rev.* **114**, 181-217 (1990).
23. Staunton, D. E., Dustin, M. L., Erikson, H. P. & Springer, T. A. *Cell* **61**, 242-254 (1990).
24. Diamond, M. S., Staunton, D. E., Marlin, S. D. & Springer, T. A. *Cell* **65**, 961-971 (1991).
25. Silvermann, L. B. *et al. J. Immunol.* **142**, 4194-4200 (1989).
26. Mentzer, S. *et al. J. exp. Med.* **165**, 1383-1392 (1987).
27. Beyers, A. D., Barclay, A. N., Law, D. A., He, Q. & Williams, A. F. *Immunol. Rev.* **111**, 59-77 (1989).
28. Nong, Y. H., Remold-O'Donnell, E., LeBien, T. W. & Remold, H. G. *J. exp. Med.* **170**, 259-267 (1989).
29. Cyster, J. G., Shotton, D. M. & Williams, A. F. *EMBO J.* **10**, 893-902 (1991).
30. Ardman, B., Sikorski, M., Settles, M. & Staunton, D. E. *J. exp. Med.* **172**, 1151-1158 (1990).
31. Gorga, J. C., Knudsen, P. J., Foran, J. A., Strominger, J. L. & Burakoff, S. J. *Cell. Immunol.* **103**, 160-173 (1986).
32. Goldstein, S. A. & Mescher, M. J. *Immunol.* **137**, 3383-3392 (1986).
33. Takai, Y., Reed, M., Burakoff, S. J. & Herrmann, S. *Proc. natn. Acad. Sci. U.S.A.* **84**, 6864-6868 (1987).

ACKNOWLEDGEMENTS. We thank Boehringer-Ingelheim for recombinant soluble ICAM-1; E. Clark for anti-ICAM LB2; J. Ledbetter for anti-CD43 G10-2 and G19-1; and R. Rothlein, B. Sleckman and S. Kourebanas for discussion and for critical review of the manuscript. J.K.P. is the recipient of a Physician-Scientist Award and a Medical-Scientist Award, W. C. Hahn is a Howard Hughes Medical Institute Predoctoral Fellow, B.E.B. is the recipient of a Clinician-Scientist Award from the American Heart Association and a McDonnell Scholar Award from the J. S. McDonnell Foundation. This work was supported by the NCI (S.J.B.) and NIAID (F.S.R.).

MHC class I surface expression in embryo-derived cell lines inducible with peptide or interferon

Elizabeth K. Bikoff*, Leah Jaffe†, Randall K. Ribaud‡, Gillis R. Otten§, Ronald N. Germain‡ & Elizabeth J. Robertson†

* Department of Obstetrics, Gynecology and Reproductive Sciences, Mount Sinai School of Medicine, New York, New York 10029, USA

† Department of Genetics and Development, Columbia University College of Physicians and Surgeons, 701 West 168th Street, New York, New York 10032, USA

‡ Laboratory of Immunology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA

It has long been recognized that the absence of expression of products of the major histocompatibility complex (MHC) during early development might allow the fetus to escape recognition by maternal lymphocytes. In addition to the MHC class I heavy chain and β_2 -microglobulin, antigenic peptide is an essential structural component of the class I molecule¹⁻⁵. Indeed, there is evidence that MHC-linked genes encoding peptide transporter molecules⁶⁻¹⁰ and possibly components of a proteolytic complex^{11,12} are necessary for MHC class I assembly and stability at the cell surface. Here we demonstrate that embryonic cells in general show a defect in MHC class I assembly. Surface expression was rescued in the presence of an appropriate antigenic peptide, or by treatment with interferon. Consistent with this, HAM1 (ref. 9) messenger RNA was not constitutively expressed, but was inducible by interferon, and during differentiation *in vitro*. Thus, tolerance of the fetal allograft may in part be controlled at the level of peptide-dependent MHC class I assembly.

We recently identified an embryonic cell line (EE2H3) that has a defect in MHC class I assembly^{13,14}. The phenotype of EE2H3 closely resembles that described for somatic cell mutants RMA-S and .174/T2 (refs 2, 3). By contrast, EE2H3 was established from primary cultures of mouse embryo cells without immunoselection, and might therefore represent a normal embryonic cell type. To test the possibility that the lack of peptide transporter function might generally be a characteristic of embryonic cells, the present study examines MHC class I assembly in a panel of H-2D^d-transfected embryo-derived cell lines (Table 1). To ensure high levels of expression of H-2D^d and β_2 -microglobulin protein, these genes were introduced under control of the human β -actin promoter.

H-2D^d-transfected EE8D2 cells do not constitutively express H-2D^d surface antigens at physiological temperature (Fig. 1). Incubation with human immunodeficiency virus (HIV) P18 peptide¹⁵ (Fig. 1a) or treatment with interferon (Fig. 1b) rescues H-2D^d surface expression. F9 EC cells do not express β_2 -microglobulin (β_2m) or conventional MHC class I transcripts¹⁶⁻¹⁸. However, induction of differentiation by retinoic acid is accompanied by the appearance of β_2m and MHC class I mRNA and surface protein^{16,18}. F9 transformants produced β_2m^b protein (Fig. 2a), and abundant levels of H-2D^d H chains reactive with anti-H-2D^d (34-2-12) monoclonal antibody directed against epitopes located in the $\alpha 3$ domain (Fig. 2b). However, little or no β_2m protein was coprecipitated with H-2D^d H chains. Moreover, these H-2D^d H chains were not recognized by 34-5-8 anti-H-2D^d antibody directed against $\alpha 1/\alpha 2$ determinants. Thus, we found no evidence for the formation of fully assembled H-2D^d molecules. Addition of peptide directly to cell lysates at 4 °C promoted correct folding of H-2D^d

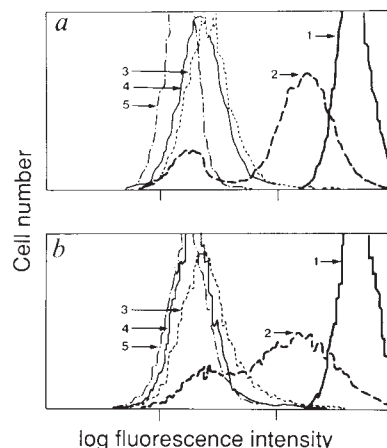


FIG. 1 Defective surface expression in H-2D^d-transfected EE8D2 cells can be rescued by the addition of an appropriate antigenic peptide or treatment with interferon. Cell-surface staining was done at 4 °C with biotin-conjugated antibody followed by fluorescein isothiocyanate (FITC)-avidin. a, DAP.H-2D^d (1), BDE 31 cells incubated overnight at 37 °C with the HIV P18 (ref. 15) peptide (50 μ M) (2) or a control env T2 peptide (50 μ M) (3), or control BDE 31 cells incubated in medium alone (4) were stained with 34-5-8 anti-H-2D^d antibody directed against $\alpha 1/\alpha 2$ epitopes. As a control, BDE 31 cells incubated overnight with the HIV P18 peptide were also stained with FITC-avidin alone (5). b, DAP.H-2D^d (1), BDE 31 cells incubated for 48 h in the presence of interferon ($\alpha + \beta$; Lee Biomolecular) (500 U ml⁻¹) (2), or control BDE 31 cells incubated in medium alone (3) were stained with 34-5-8 anti-H-2D^d antibody directed against $\alpha 1/\alpha 2$ epitopes. As a control, BDE 31 cells treated with interferon were stained with B22-249 anti-H-2D^d antibody (4) or FITC-avidin alone (5).

§ Present address: Cell Genesys, 344 Lakeside Drive, Foster City, California 94404, USA.

TABLE 1 Cell lines used in these experiments

Cell line	Description	Reference	Gene constructs	Transformants
DAP-3	Thymidine kinase negative L cell subline	24	HuActH-2D ^d	DAP.H-2D ^d
EE2H3	Src-transformed fibroblast-like cell line isolated from C3H/HeJ (H-2 ^k) day 11 embryos	13	HuActH-2D ^d	EE2H3.H-2D ^d
EE8D2	HaSV-transformed fibroblast-like cell line isolated from C3H/HeJ (H-2 ^k) day 8 embryos	13	[HuActH-2D ^d +HuAct β 2m]	BDE #31
EE2G6	A-MuLV-transformed fibroblast-like cell line isolated from C3H/HeJ(H-2 ^k) day 11 embryos	13	—	—
F9	Embryonal carcinoma (EC) cell line derived from a 129/Sv (H-2 ^b) teratocarcinoma	25	[HuActH-2D ^d +HuAct β 2m]	BDF #7, #12 and #135
CCE	Pluripotent embryonic stem (ES) cell line derived from a 129/Sv (H-2 ^b) blastocyst stage embryo	26	HuActH-2D ^d	CCE.H-2D ^d

A construct (designated LKD1) for expression of the *H-2D^d* gene under control of the human β -actin promoter was previously described¹⁴. Briefly, a 5.1-kilobase (kb) *Bam*H1 fragment containing the translational start site, the coding region of the *H-2D^d* gene plus the polyadenylation signal was inserted into LK444 (ref. 27), an expression vector consisting of the human β -actin gene 5' flanking sequence + 5'UT region + IVS 1, followed by the 3' splice acceptor site, a short polylinker containing unique *Sal*1, *Hind*III and *Bam*H1 restriction sites, a SV40 poly(A) signal and the G418^R gene from pSV2neo. DAP3.H-2D^d (previously designated LKD13) strongly expresses H-2D^d surface protein. The EE2H3.H-2D^d transformant (previously designated LKD8) showing a defect in H-2D^d assembly has been described¹⁴. The HuActH-2D^d construct used here was similar except that the 3' *Eco*R1 site contributed by the original pD^d-1 genomic clone was left intact. This allowed us to recover a 9.4-kb *Eco*R1 fragment comprising the human β -actin promoter regulatory sequences and the *H-2D^d* gene (HuActH-2D^d) that could be cotransfected along with a plasmid designated HuAct β 2m, that directs high amounts of expression of β 2m under control of the human β -actin promoter. Briefly, a synthetic oligonucleotide comprising *Eco*R1-*Xba*1-*Kpn*1-*Xba*1-*Eco*R1 recognition sites was introduced into the unique *Eco*R1 site present in LK444, and the resultant plasmid designated LKXX. A full-length complementary DNA clone encoding the C57BL/6 allelic form of β 2m (that is, β 2m^b) having *Sal*1/*Sma*1 ends was inserted into *Bam*H1-digested, blunt-ended, *Sal*1-digested LKXX to yield the HuAct β 2m construct. The 9.4-kb HuActH-2D^d *Eco*R1 fragment (12 μ g) + 3 μ g *Xba*1-digested HuAct β 2m DNA was introduced into EE8D2 or F9 cells by calcium phosphate precipitation, and transformants selected in the presence of G418. On the basis of the observation that so-called 'empty' class I molecules can be stabilized at 25 °C²⁸, transformants were screened after overnight culture at room temperature. Clones showing serologically detectable surface H-2D^d protein induced by cooling were expanded for further study. BDE 31 is one of several [HuActH-2D^d+HuAct β 2m] $\times$ EE8D2 transfected cell lines producing cytoplasmic H-2D^d+ β 2m protein. LKD1 plasmid DNA was introduced into CCE ES cells by electroporation, and individual G418^R colonies screened for surface H-2D^d expression after differentiation *in vitro*.

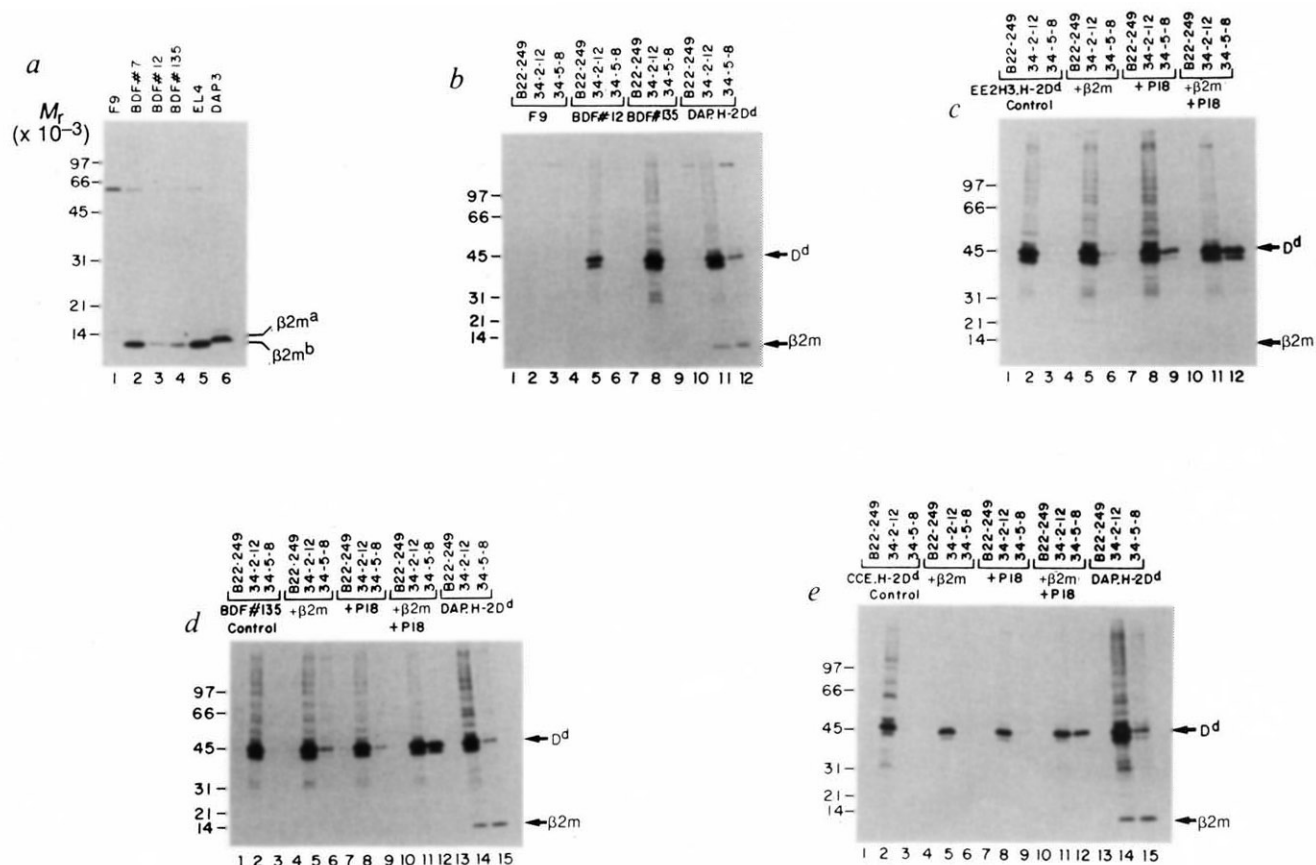


FIG. 2 H-2D^d chains and β 2m protein produced by embryonic cell lines. Cells were biosynthetically labelled with ³⁵S-methionine for 40 min. Cytoplasmic extracts were immunoprecipitated using the indicated antibodies and analysed by SDS-PAGE under reducing conditions. The arrows indicate the positions of the β 2m protein and the 43K H-2D^d H chains. Lysates prepared from the indicated cell lines were precipitated with a rabbit anti-human β 2m

antibody (Boehringer Mannheim) (a) or monoclonal anti-MHC class I antibodies as indicated (b). The smaller band recognized by 34-2-12 anti-H-2D^d antibody directed against α 3 determinants represents immature H-2D^d chains. Lysates prepared from EE2H3.H-2D^d (c) BDF 135 (d) or CCE.H-2D^d (e) were precleared overnight at 4 °C in the presence of HIV P18 peptide (50 μ M) and/or human β 2m (1.7 μ M) (CalBiochem) as indicated.

H chains (Fig. 2c-e). We consistently observed a large induction of $\alpha 1/\alpha 2$ epitopes on addition of both peptide and $\beta 2m$. This is consistent with the actions of peptide and $\beta 2m$ in lysates of RMA-S and .174 cells⁴. The inability to express $\alpha 1/\alpha 2$ epitopes cannot simply be due to inefficient expression of $\beta 2m$, because even in the presence of excess $\beta 2m$, peptide was necessary for this conformational change. Peptide-dependent MHC class I assembly was observed with the entire panel of embryo-derived cell lines.

Cell-surface staining experiments confirmed that H-2D^d protein was weakly expressed at the cell surface. Addition of peptide markedly enhanced H-2D^d surface expression (Fig. 3a, b). Peptide was necessary even in the presence of excess $\beta 2m$

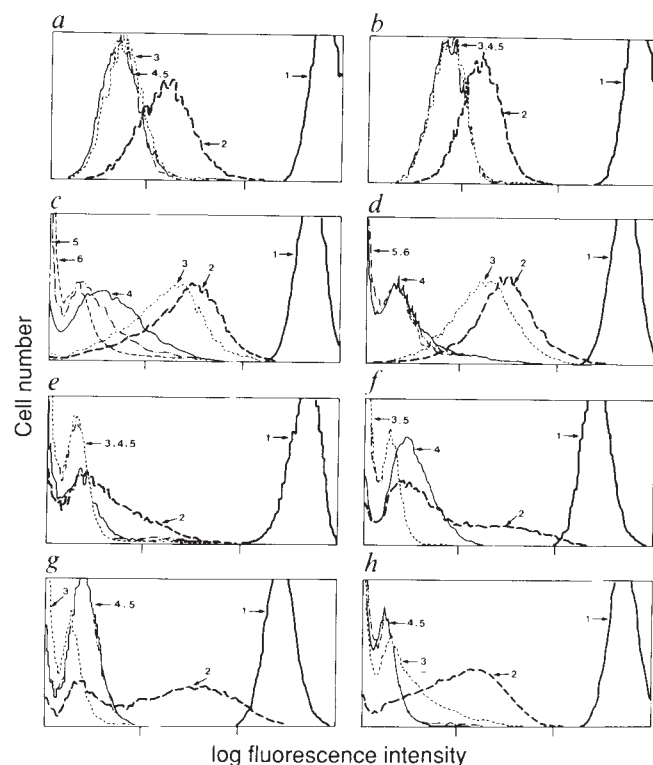


FIG. 3 Induction of surface expression in H-2D^d-transfected F9 cells. Cell-surface staining was done at 4 °C with biotin-conjugated antibody followed by FITC-avidin. **a**, DAP.H-2D^d (1), BDF 12 cells incubated overnight at 37 °C with the HIV P18 peptide (50 μ M) (2) or a control env T2 peptide (50 μ M) (3), or control BDF 12 cells incubated in medium alone (4) were stained with 34-5-8 anti-H-2D^d antibody directed against $\alpha 1/\alpha 2$ epitopes. As a control, BDF 12 cells incubated overnight with the HIV P18 peptide were stained with B22-249 anti-H-2D^b antibody (5). **b**, The same experiment as shown in **a**, except with BDF 135 cells. **c**, DAP.H-2D^d (1), BDF 12 cells incubated overnight at 37 °C with HIV P18 (50 μ M) + Hu $\beta 2m$ (1 μ M) (2), HIV P18 (3), or Hu $\beta 2m$ alone (4), or control BDF 12 cells incubated in medium alone (5) were stained with 34-5-8 anti-H-2D^d antibody. As a control, BDF 12 cells incubated overnight with HIV P18 + Hu $\beta 2m$ were stained with FITC-avidin alone (6). **d**, The same experiment as shown in **c**, except with BDF 135 cells. **e**, EL4 (1), F9 cells treated with retinoic acid (2), or control F9 cells incubated in medium alone (3) were stained with B22-249 anti-H-2D^b antibody. As a control, retinoic acid-treated F9 (4) or control F9 cells (5) were also stained with FITC-avidin alone. **f**, DAP.H-2D^d (1), BDF 7 cells treated with retinoic acid (2), or control BDF 7 cells incubated in medium alone (3) were stained with 34-5-8 anti-H-2D^d antibody. As a control, retinoic acid-treated BDF 7 (4) or control BDF 7 (5) were stained with FITC-avidin alone. **g**, The same experiment as shown in **f**, except with BDF 135 cells. **h**, DAP.H-2D^d (1), BDF 12 cells treated with interferon (2), or control BDF 12 cells incubated in medium alone (3) were stained with 34-5-8 anti-H-2D^d antibody. As a control, BDF 12 cells treated with interferon (4) or control BDF 12 cells (5) were also stained with FITC-avidin alone. To test the requirement for peptide in the presence of excess $\beta 2m$, all these experiments were done with cell populations cultured in the presence of 10% FCS containing high amounts of bovine $\beta 2m$.

(Fig. 3c, d). As expected, we observed a large induction of H-2D^d surface protein when F9 transformants were induced to differentiate after treatment with retinoic acid (Fig. 3f, g). Surface H-2D^d was greatly increased in response to treatment with interferon (Fig. 3h).

Interferon induces expression of an ATP-dependent peptide transporter⁷ that participates in MHC class I assembly¹⁰. Homologous genes in the class II regions of the rat, human and mouse MHC, designated Mtp-1 (ref. 6), Ring-4 (refs 7, 8) and HAM1 (ref. 9), respectively, encode proteins related to the ABC superfamily of transporters. It was therefore of interest to test whether defective MHC class I assembly in embryonic cells was associated with a lack of expression of HAM1 mRNA. This was indeed the case. HAM1 mRNA expression was inducible by treatment with interferon (Fig. 4a), or during differentiation *in vitro* (Fig. 4b). Thus, there was a good correlation between expression of surface H-2D^d molecules and expression of the HAM1 transporter.

These data show that HAM1 transcription is developmentally regulated. Future experiments will examine HAM1 expression in normal embryos. Although the onset of $\beta 2m$ and MHC class I H-chain transcription has been detected as early as day 7.5 post coitum^{19,20}, surface protein was not observed until day 10.5

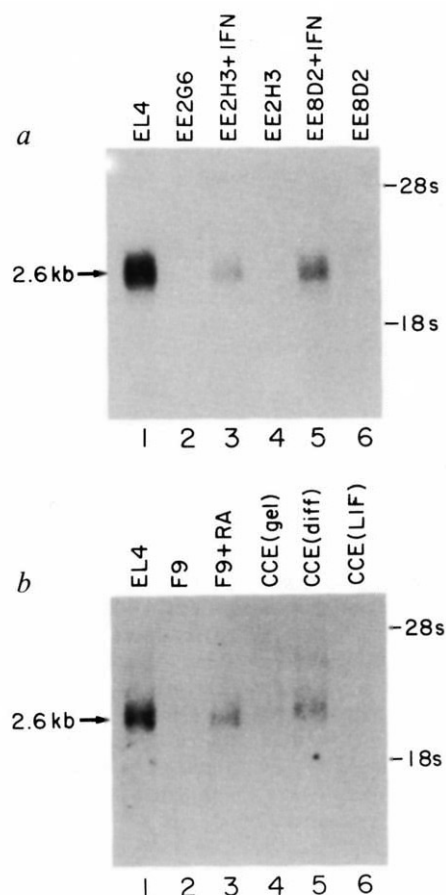


FIG. 4 HAM1 mRNA expression in embryo-derived cell lines. Northern blot analysis of total RNA (30 μ g per lane) from the indicated cell lines. HAM1 transcripts were detected with an EcoRI fragment from the rat mtp-1 cDNA clone 510-15 (kindly provided by J. Howard and G. Butcher). **a**, Where indicated, interferon (500 units ml⁻¹) was added 48 h before RNA extraction. **b**, ES cells were induced to differentiate *in vitro*²⁹ or F9 cells treated with retinoic acid³⁰ as described. The positions of 18S and 28S ribosomal RNA are indicated.

METHODS. RNA was size-fractionated on formaldehyde agarose gels, transferred to GeneScreen Plus, and crosslinked by ultraviolet irradiation. Filters were hybridized overnight at 45 °C, and washed for 60 min at 65 °C in 2 $\times$ SSC + 1% SDS. A nick-translated 1.4-kb EcoRI fragment from the 3' portion of the rat Mtp-1 510-15 cDNA was used as the probe.

post coitum²¹. At this stage, only a small percentage of cells recovered from embryos express detectable class I surface antigens²¹. We recently identified restricted cell populations in early postimplantation stage embryos that express MHC class I and $\beta 2m$ transcripts^{22,23}. It will be important to test whether these fetal tissues express all the components of the MHC class I peptide-loading machinery. □

Received 25 July; accepted 26 September 1991.

1. Bjorkman, P. J. *et al.* *Nature* **329**, 506–512 (1987).
2. Townsend, A. *et al.* *Nature* **340**, 443–448 (1989).
3. Cerundolo, V. *et al.* *Nature* **345**, 449–452 (1990).
4. Townsend, A. *et al.* *Cell* **62**, 285–295 (1990).
5. Silver, M. L., Parker, K. C. & Wiley, D. C. *Nature* **350**, 619–622 (1991).
6. Deverson, E. V. *et al.* *Nature* **348**, 738–741 (1990).
7. Trowsdale, J. *et al.* *Nature* **348**, 741–744 (1991).
8. Spies, T. *et al.* *Nature* **348**, 744–747 (1991).
9. Monaco, J. J., Cho, S. & Attaya, M. *Science* **250**, 1723–1726 (1990).
10. Spies, T. & DeMars, R. *Nature* **351**, 323–324 (1991).
11. Monaco, J. J. & McDevitt, H. O. *Proc. natn. Acad. Sci. U.S.A.* **79**, 3001–3005 (1982).
12. Monaco, J. J. & McDevitt, H. O. *Nature* **309**, 797–799 (1984).
13. Silverman, T. *et al.* *J. Immun.* **140**, 4378–4387 (1988).
14. Bikoff, E. K., Otten, G. R. & Robertson, E. J. *Eur. J. Immun.* **21**, 1997–2004 (1991).

15. Takahashi, H. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **85**, 3105–3109 (1988).
16. Croce, C. M. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **78**, 5754–5758 (1981).
17. Morello, D. *et al.* *Nature* **296**, 260–262 (1982).
18. Rosenthal, A., Wright, S., Cedar, H., Flavell, R. & Grosveld, F. *Nature* **310**, 415–418 (1984).
19. Morello, D., Duprey, P., Israel, A. & Babinet, C. *Immunogenetics* **22**, 441–452 (1985).
20. Hedley, M. L., Drake, B. L., Head, J. R., Tucker, P. W. & Forman, J. J. *Immun.* **142**, 4046–4053 (1989).
21. Ozato, K., Wan, Y. J. & Orrison, B. M. *Proc. natn. Acad. Sci. U.S.A.* **82**, 2427–2431 (1985).
22. Jaffe, L., Jeannotte, L., Bikoff, E. K. & Robertson, E. J. *J. Immun.* **145**, 3474–3482 (1990).
23. Jaffe, L., Robertson, E. J. & Bikoff, E. K. *J. Immun.* (in the press).
24. Margulies, D. H. *et al.* *J. Immun.* **130**, 463–470 (1983).
25. Bernstine, E. G., Hooper, M. L., Grandchamp, S. & Ephrussi, B. *Proc. natn. Acad. Sci. U.S.A.* **70**, 3899–3903 (1973).
26. Robertson, E., Bradley, A., Kuehn, M. & Evans, M. *Nature* **323**, 445–448 (1986).
27. Gunning, P., Leavitt, J., Muscat, G., Ng, S. Y. & Kedes, L. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4831–4835 (1987).
28. Ljunggren, H. G. *et al.* *Nature* **346**, 476–480 (1990).
29. Robertson, E. J. in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach* (ed. Robertson, E. J.) 71–112 (IRL, Oxford, 1987).
30. Rudnicki, M. A. & McBurney, M. W. in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach* (ed. Robertson, E. J.) 19–49 (IRL, Oxford, 1987).

ACKNOWLEDGEMENTS. We thank A. Townsend for valuable comments on the manuscript, D. Margulies for discussions, E. Okoye and Y. Duzant for excellent technical assistance, K. Ozato for cell lines, J. Berzofsky for the HIV P18 and env T2 peptides, J. Howard and G. Butcher for the Mtp-1 cDNA clone, and A. Pizzimenti for assistance with flow cytometry. This work was supported by a grant from the NIH to E.K.B. and E.J.R.; G.R.O. was the recipient of a fellowship from the American Cancer Society; R.K.R. was supported by the Cancer Research Institute and the National Multiple Sclerosis Society; E.J.R. is the recipient of a Leukemia Society Scholarship.

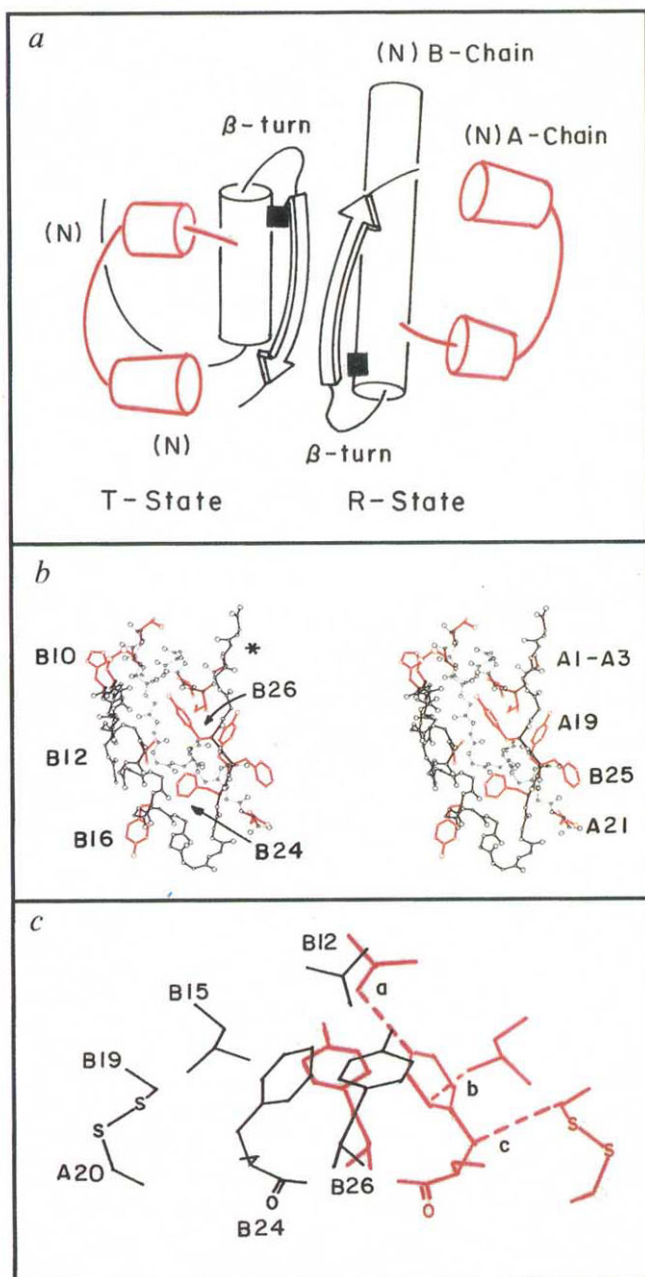
Receptor binding redefined by a structural switch in a mutant human insulin

Qing Xin Hua^{*†}, Steven E. Shoelson[‡],
Michel Kochoyan^{*†} & Michael A. Weiss^{*§||}

^{*} Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA
[‡] Joslin Diabetes Center and Departments of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02215, USA
[§] Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts 02114, USA

CRYSTAL structures of insulin have been determined in various distinct forms^{1–5}, the relevance of which to receptor recognition has long been the subject of speculation^{2,6,7}. Recently the crystal structure of an inactive insulin analogue has been determined and, surprisingly, found to have a conformation identical to native insulin^{8,9}. On this basis Dodson and colleagues have suggested that the known insulin crystal structures reflect an inactive conformation, and that a change in conformation is required for activity—specifically, the carboxy terminal residues of the B-chain are proposed to separate from the amino terminal residues of the A-chain⁸. Here we report the solution structure of an active insulin mutant^{7,10}, determined by two-dimensional NMR, which supports this hypothesis. In the mutant, the carboxy terminal β -turn and β -strand of the B-chain are destabilized and do not pack across the rest of the molecule. We suggest that analogous detachment of the carboxy terminal region of the B-chain occurs in native insulin on binding to its receptor. Our finding that partial unfolding of the B-chain exposes an alternative protein surface rationalizes the receptor-binding properties of a series of anomalous insulin analogues^{7,11–13}, including a mutant insulin associated with diabetes mellitus in man^{14,15}.

Insulin consists of two chains, the A-chain (21 residues) and B-chain (30 residues)¹⁶, and binds to its receptor as a monomer¹⁷. In crystallographic dimers and hexamers^{2–5} two families of structures are observed, the T and R states⁵ (Fig. 1a). In the T state the A-chain contains N- and C-terminal α -helices (residues A1–A8 and A12–A18); the B-chain contains an extended



[†] Permanent addresses: Institute of Biophysics, Academia Sinica, Beijing, China (Q.X.H.); CNRS URA 32, Biop. Polytechnique, 91128 Palaiseau, France (M.K.).

^{||} To whom correspondence should be addressed at Harvard Medical School.

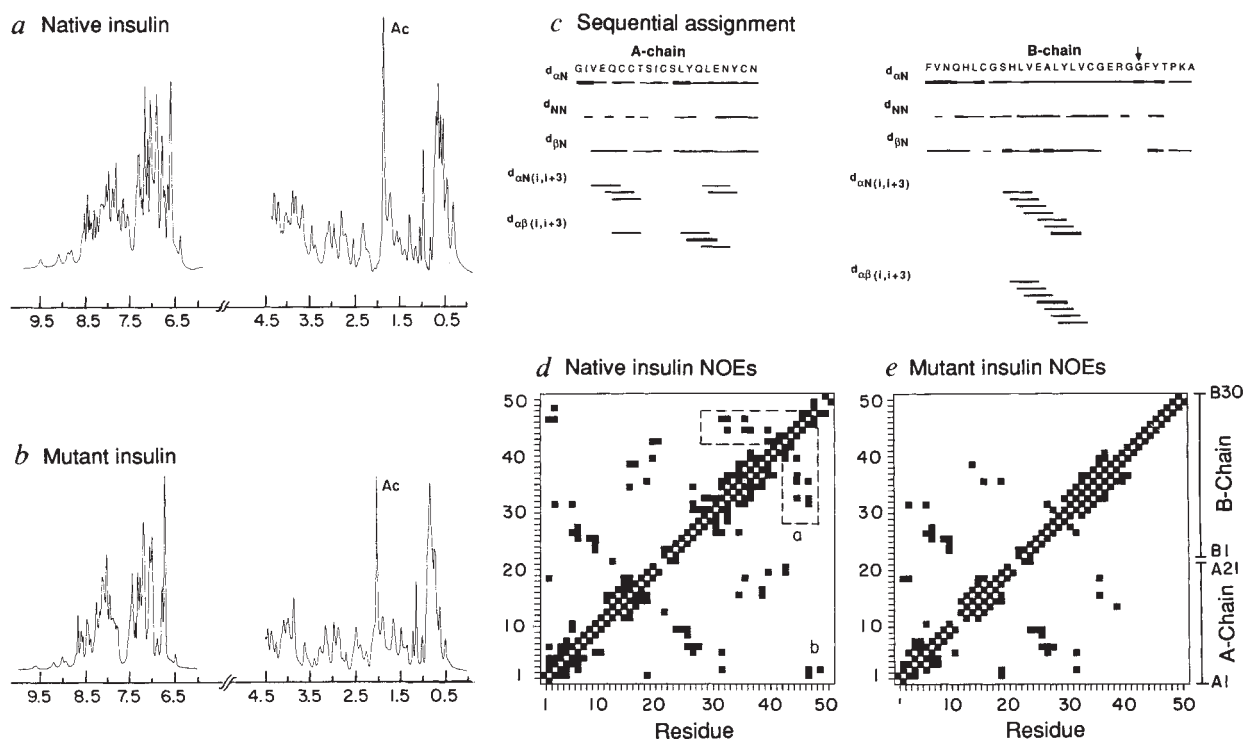


FIG. 2 *a* and *b*, ^1H -NMR spectra of *a*, native monomeric (1.0 mM) and *b*, mutant insulin (1.5 mM) in 80% H_2O -20% deuterated acetic acid at 25 $^\circ\text{C}$. Two-dimensional (2D) NMR analysis¹⁹ is similar to that of engineered monomer under physiological conditions²⁶. Ac indicates a residual resonance of deuterated acetic acid. In *b*, the ^1H -NMR spectrum of Gly B24-insulin is shown under similar conditions²⁷. Residues B20-B30 exhibit motional narrowing and a more limited range of chemical shifts. The latter are similar to those of a synthetic B23-B30 octapeptide, which is disordered in solution²⁷. Unlike native insulin, Gly B24 insulin does not dimerize in aqueous solution (pH 1.9); its NMR features are similar in 0-20% acetic acid²⁷. *c*,

Summary of sequential and medium-range NOE connectivities²¹ in NOESY spectrum of Gly B24 insulin²⁷. Sequential assignments of insulin or analogues have been described by others²⁸⁻³⁰. *d*, Summary of inter-residue NOE contacts in native insulin. Residue numbers 1-21 refer to A-chain; 22-51 refer to B-chain (B1-B30). NOE contacts *a* (boxed) orient B-chain β -strand (B24-B28) against B-chain α -helix; NOE contacts *b* orient B-chain β -strand against the N-terminal α -helix of A-chain. Not shown: disulphide bonds A6-A11, A7-B7 and A20-B19. *e*, Summary of NOE contacts in Gly B24 insulin. Absence of B24-B26 NOEs are verified by ^{13}C -edited NOESY studies of selectively labelled protein²⁷.

N-terminal arm, central α -helix (B9-B19), β -turn (B20-B23), and C-terminal β -strand (B24-B28). The R state, an allosteric feature of hexamer assembly⁵, exhibits an additional α -helical transition (B1-B9). The physiological relevance of T and R structures has not been established. The receptor-binding affinity⁹ of mini-proinsulin, which forms an isomorphous T_3R_3 hexamer⁸, is reduced by at least 10^6 .

Phenylalanine at the B24 position (Phe B24) is invariant among insulin sequences¹⁸. In crystal structures the Phe B24 side chain makes tertiary contacts with other conserved residues (Fig. 1*b* and *c*) that stabilize the orientation of the B-chain β -turn (B20-B23) and β -strand (B24-B28) to define a compact hydrophobic core²⁻⁵. The importance of Phe B24 to bioactivity⁷ has been inferred from studies of low-potency B24 analogues⁷.

Remarkably, however, certain B24 substitutions are well tolerated, including glycine and D-amino acids^{7,10,11}. These analogues are regarded as 'anomalous' as their high bioactivity cannot readily be explained by crystal models^{2,8}. The receptor-binding affinity of Gly B24-insulin (a complete agonist) is 78% of that of native insulin⁷.

^1H -NMR spectra of the native and mutant insulins are shown in Fig. 2*a* and *b*. Sequential resonance assignment, analysis of secondary structure, and long-range nuclear Overhauser enhancements (NOEs; Fig. 2*d*) of native insulin are in accord with the T state¹⁹. Gly B24 insulin (Fig. 2*c*) shows that it has similar (but not identical) α -helices (A3-A8, A13-A18 and B9-B19). Unlike native insulin, no evidence is obtained for a β -turn in the B-chain (residues B20-B23). Inter-residue NOEs in Gly B24 insulin are summarized in Fig. 2*e*. Long-range interactions between the A-chain and B-chain residues B1-B19 persist, demonstrating that a significant portion of the insulin core is retained. Absent, however, are long-range NOEs between the B-chain β -strand (B24-B28) and the A- and B-chain α -helices; in native insulin these NOEs define the orientation of the β -strand (Fig. 2*d*). NOE contacts thus altered in Gly B24 insulin include residues (A2, A3, B12, B15, B16, B25 and B26) that are proposed to be critical in folding and receptor recognition (Fig. 1*b*)^{2,6,20}.

Calculations to determine the tertiary structure of Gly B24 insulin and native human insulin were made using distance-geometry and restrained molecular dynamics^{21,22}. The backbone configuration of a representative solution of native insulin is shown in Fig. 3*a* (central panel). The orientation of the two chains is similar to that of a representative T-state crystal

FIG. 1 *a*, Cylinder representation of 4-Zn (T_3R_3) crystallographic dimer³ showing T and R states: A-chains (red) and B-chain (black). B-chain residues B20-B23 form a β -turn; B24-B28 form a dimer-related antiparallel β -sheet. The position of Phe B24 is indicated by a filled square. *b*, Crystallographic protomer (stereo) in the T-state (molecule 2 or 2-Zn crystal²): B-chain (heavy line) and A-chain (thin line). Side chains required for bioactivity²⁰ or stability in the crystal² are shown in red (adapted from ref. 2). Asterisk: site of B29-A1 crosslink in mini-proinsulin^{8,9}. *c*, Interactions of Phe B24 in the 2-Zn crystallographic dimer²: molecule 1 (red) and molecule 2 (black). In the protomer Phe B24 contacts the B-chain α -helix at Val B12 (*a*), Leu B15 (*b*) and Cys B19 (*c*). Other contacts in the dimer are shown. Intramolecular contacts are retained in the crystal structure of the monomeric analogue DPI^{23,24}. Contacts *a* and *b* are observed in the NOESY spectrum of native insulin monomer¹⁹; NOE of *c* is not resolved but is consistent with distance-geometry/restrained molecular dynamics (DG/RMD) structures (Fig. 3).

structure² (left-hand panel). A corresponding model of Gly B24 insulin is shown in the right-hand panel. The α -helical core is retained; strikingly, the B-chain β -turn and C-terminal β -strand are disordered. Ensembles of backbone structures are shown in Fig. 3b for human insulin (central panel) and for Gly B24 insulin (right-hand panel). The structures are aligned along the B-chain α -helix (residues B9–B19) as in an earlier analysis of crystallographic protomers¹ (left-hand panel of Fig. 3b).

The structure of the C-terminal region of the B-chain is stabilized in crystals by dimerization (Fig. 1a). That this structure may not represent the active form of the insulin monomer has been suggested by a puzzling pattern of receptor-binding properties among B-chain analogues^{7,10–12}; loss or gain of bioac-

tivity is proposed to correspond to the 'locking' (inactive) or 'unlocking' (active) of productive conformations. We propose that the NMR structure of Gly B24 insulin represents a model of the unlocked state. This proposal extends an interpretation of crystallographic studies of the truncated analogue despen-tapeptide (B26–B30) insulin (DPI)^{23,24} and is supported by the high potency of such analogues and others containing destabilizing mutations in the B-chain β -turn²⁵. Why the β -turn/ β -strand structure is apparently conserved (including Phe B24) is unclear and may relate to other biological constraints, such as disulphide pairing, prohormone processing or storage.

Detachment of the B-chain β -strand reorganizes the protein surface, exposing side chains that are highly conserved (Ile or

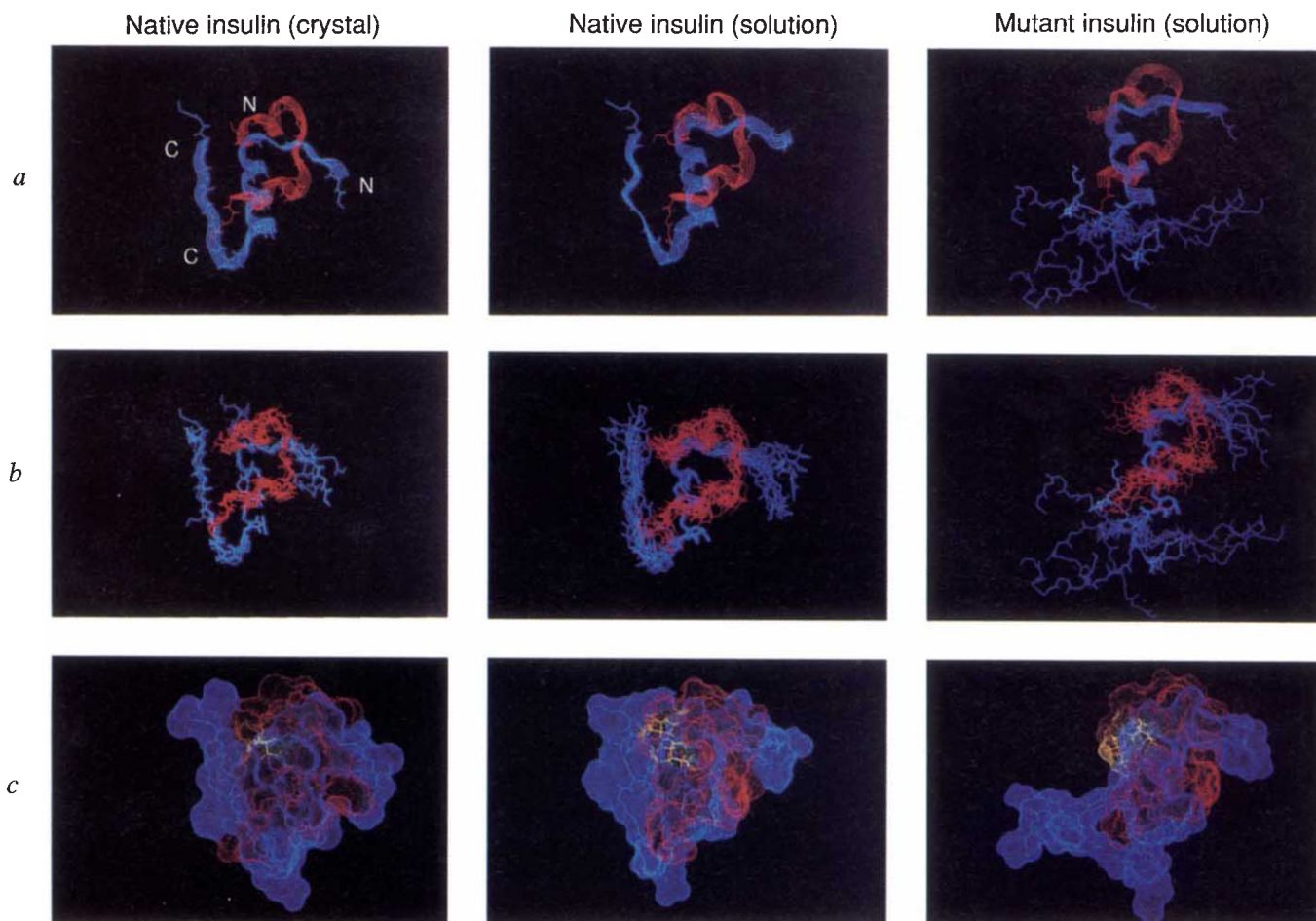


FIG. 3 *a*, Ribbon models of 2-Zn crystal structure of porcine insulin² (left), DG/RMD structure of native human insulin (centre), and DG/RMD structure of Gly B24 insulin (right). Residues B20–B30 are shown as an ensemble of backbone configurations based on random sampling of conformational space in absence of restraints; loss of ordered structure is independently verified by circular dichroism^{10,27}. Colour code: A-chain (red), B-chain (blue). *b*, Left, crystallographic protomers^{2–5} aligned in reference to B-chain α -helix (main chain only)¹; centre, ensemble of 11 NMR-derived^{21,22} solution structures of native insulin similarly aligned (backbone only); right, 11 NMR structures of Gly B24 insulin. Residues B20–B30 are disordered. Colour code as in *a*. *c*, Left, solvent-accessible surface of crystal structure (2-Zn molecule 2; ref. 2). The surface contributed by Ile A2 and Val A3 is shown in yellow, otherwise the A-chain is shown in red and the B-chain in blue; centre, solvent-accessible surface of representative DG/RMD structure of native insulin (same colour scheme); right, solvent-accessible surface of a DG/RMD structure of Gly B24 insulin. Displacement of residues B20–B30 results in significant exposure to solvent of residues Ile A2 and Val A3.

METHODS. NMR restraints for native insulin: 422 NOEs (160 sequential, 86 medium and 176 long-range) and 43 J -coupling (dihedral angle: 28 ϕ and 15 χ 1). The number of NOEs per residue (less than that expected of globular proteins) reflects the experimental data, which contain fewer NOEs than predicted by relaxation-matrix calculations based on any one crystal or DG

structure^{18,24}, presumably as a result of internal motions. NOEs (mixing times, 100 and 200 ms) were classified as strong (<2.7 Å), medium (<3.3 Å) or weak (<4.3 Å) according to the relative amplitudes of two-dimensional crosspeaks relative to tyrosine (2.5 Å) and histidine (4.3 Å) ring crosspeaks. Distance-bound corrections were made for methyl groups, degenerate aromatic ring resonances, and methylene protons without stereospecific assignments. Dihedral angle restraints were introduced based on J -coupling constants in DPI and native insulin^{21,22}: small $^3J_{\text{NH}\alpha}$ coupling constants (<5.5 Hz following linewidth correction) were constrained between -90 to -40° ; those with large $^3J_{\text{NH}\alpha}$ coupling constants (>8.5 Hz) in the absence of significant line broadening were constrained between -160 and -80° . Restraints were not introduced based on intermediate J values. χ 1 dihedral angles (residues A5–A7, A10, A12, A14, A19, B4–B6, B10, B11, B19 and B24) were restrained with tolerance of $\pm 40^\circ$. Restraints for Gly B24 insulin: 290 NOEs (135 sequential, 57 medium and 98 long-range) and 41 J -coupling (dihedral angle: 27 ϕ and 14 χ 1). DG calculations²¹ were performed with program DG-II (T. F. Havel, Harvard Medical School). RMD calculations²² were performed with program XPLOR (A. Brünger, Yale University); NOE and dihedral-angle restraints were introduced with force constants of 40 kcal per Å² and 40 kcal per radian², respectively. Statistical parameters of ensembles are given in Table 1.

TABLE 1 Statistical parameters describing NMR ensembles

	Native insulin	Mutant insulin
Root-mean-square deviations (excluding disordered regions*)		
Main chain		
A-chain	0.73 Å	0.76 Å
B-chain	1.22 Å	1.68 Å
α -helices†	0.69 Å	0.62 Å
Side chains		
A-chain	1.69 Å	1.68 Å
B-chain	2.23 Å	1.49 Å
α -helices	1.56 Å	1.48 Å
Average NOE restraint violations‡		
Upper-bound violations	0.032 Å	0.024 Å
Deviations from ideal covalent geometry		
Bond lengths	0.011 Å	0.010 Å
Bond angles	2.98°	2.73°
Restrained empirical energy function§		
Total	161.5 kcal mol ⁻¹	84.3 kcal mol ⁻¹
Constrained dihedral angles	1.2	0.9
NOE restraint energy	18.0	7.41
van der Waals	-218.2	-216.9
Hydrogen bonds	-31.2	-29.1
Improper dihedral angles	4.2	2.7
Dihedral angles	190.8	163.2
Covalent bond lengths	14.9	11.9
Bond angles	183.1	155.2

* Ensembles were aligned according to C α positions of stably folded regions (native insulin, A2-A19 and B4-B28; mutant insulin, A2-A19 and B4-B19). Root-mean-square values correspond to these regions.

† α -Helices as defined by NMR of native insulin include A2-A8, A12-A18 and B9-B19; α -helices in Gly B24 insulin are A3-A8, A13-A18 and B9-B19.

‡ Restraints used in calculating NMR ensembles are described in the legend to Fig. 3.

§ Empirical energies were calculated with the CHARMM potential energy function as described in the legend to Fig. 3.

Val A2) or invariant (Val A3)¹⁸ in the N-terminal α -helix of the A-chain. The different environments of these side chains in the locked and unlocked states are shown in Fig. 3c. In the open state—but not the closed—these hydrophobic side chains would be accessible for direct contact with the insulin receptor. Accordingly, we suggest that their strict conservation (that is, the apparent exclusion of other hydrophobic side chains) reflects the stereochemistry of the insulin-receptor contact rather than internal packing constraints as inferred from crystal structures. This model in turn predicts that the diabetes-associated mutation Val A3 $\rightarrow$ Leu (refs 14, 15) may, as a seeming paradox, be structurally conservative (in the closed state), despite its significant loss of function. Additional diabetes-associated mutations occur at positions B24 and B25 (ref. 14), emphasizing the importance of this region in receptor binding⁷. Although residues B20-B30 are unfolded in Gly B24 insulin, the present study provides no information regarding their structure or dynamics in the receptor-bound state.

In summary, crystallographic studies of insulin in distinct forms have provided a model for analysis of conformational change in proteins¹. To test the relevance of these forms to receptor recognition, we have determined the two-dimensional NMR structures of native insulin and a mutant whose high bioactivity is difficult to rationalize from crystal structures. Although the solution structure of native insulin retains overall features of the crystallographic T state, Gly B24-insulin exhibits partial unfolding of the B-chain with detachment of A-chain/B-chain contacts. These results have direct implications for the mechanism of insulin action and for the rational design of insulin agonists. The striking complementarity of NMR and crystallographic studies⁸ of open and closed analogues has broad relevance to understanding the inter-relationship of protein function, structure and dynamics. □

Received 7 August; accepted 3 October 1991.

- Chothia, C., Lesk, A. M., Dodson, G. G. & Hodgkin, D. C. *Nature* **302**, 500-505 (1983).
- Baker, E. N. *et al. Phil. Trans. R. Soc. B319*, 389-456 (1988).
- Bentley, G., Dodson, E., Dodson, G., Hodgkin, D. & Mercola, D. *Nature* **261**, 166-168 (1976).
- Badger, J. *et al. Acta crystallogr. B47*, 127-136 (1991).
- Derewenda, U. *et al. Nature* **338**, 594-596 (1989).
- Dodson, G. G., Hubbard, R. E. & Reynolds, C. D. *Biopolymers* **22**, 281-292 (1983).
- Mirmira, R. & Tager, H. S. *J. biol. Chem.* **264**, 6349-6354 (1989).
- Derewenda, U. *et al. J. molec. Biol.* **220**, 425-433 (1991).
- Markussen, J. *Int. J. Pept. Protein Res.* **26**, 70-77 (1985).
- Shoelson, S. E., Lu, Z.-X., Parlautan, L., Lynch, C. S. & Weiss, M. A. *Biochemistry* (submitted).
- Kobayashi, M. *et al. Biochem. biophys. Res. Commun.* **107**, 329-336 (1982).
- Nakagawa, S. H. & Tager, H. S. *J. biol. Chem.* **261**, 7332-7341 (1986).
- Casaretto, M. *et al. Biol. Chem. Hoppe-Seyler* **368**, 709-716 (1987).
- Shoelson, S. E. *et al. Nature* **302**, 540-542 (1983).
- Nanjio, A. C. T. *et al. J. clin. Invest.* **77**, 514-519 (1986).
- Brown, H., Sanger, F. & Kitai, R. *Biochem. J.* **60**, 556-565 (1955).
- Brange, J. *et al. Nature* **333**, 679-682 (1988).
- Steiner, D. F., Bell, G. I. & Tager, H. S. in *Endocrinology* (ed. DeGroot, L. J.) 1263-1289 (Saunders, Philadelphia, 1990).
- Hua, Q. X. & Weiss, M. A. *Biochemistry* **30**, 5505-5515 (1991).
- Pullen, R. A. *et al. Nature* **259**, 369-373 (1976).
- Wuthrich, K. *Meth. Enzym.* **177**, 125-136 (1989).
- Clare, G. M. & Gronenborn, A. M. *Crit. Rev. Biochem. molec. Biol.* **24**, 479-563 (1989).
- Bi, R. C. *et al. Biopolymers* **23**, 391-395 (1984).
- Dai, J.-B., Lou, M.-Z., You, J.-M. & Liang, D.-C. *Sci. China* **30**, 55-65 (1987).
- Wang, S.-h., Hu, S.-q., Burke, G. T. & Katsoyannis, P. J. *Protein Chem.* **10**, 313-324 (1991).
- Weiss, M. A., Hua, Q. X., Lynch, C. L., Frank, B. H. & Shoelson, S. E. *Biochemistry* **30**, 7373-7389 (1991).
- Hua, Q. X., Shoelson, S. E. & Weiss, M. A. *J. molec. Biol.* (submitted).
- Kline, A. D. & Justice, R. M. *Jr Biochemistry* **29**, 2906-2913 (1990).
- Boelens, R., Ganaou, M. L., Verheyden, P. & Kaptein, R. *Eur. J. Biochem.* **191**, 147-153 (1990).
- Kristensen, S. M., Jorgensen, A. M., Led, J. J., Balschmidt, P. & Hansen, F. B. *J. molec. Biol.* **218**, 221-231 (1991).

ACKNOWLEDGEMENTS. This work was supported in part by grants from NIH (M.A.W.), NSF (S.E.S.), American Diabetes Association, and Juvenile Diabetes Foundation International (M.A.W. and S.E.S.). NMR spectra were obtained at Harvard Medical School (NIH 1 S10 RR04862-01). We thank A. Jasanoff for photography; W. Jia for technical assistance; T. L. Blundell, R. DiMarchi, G. G. Dodson, M. Karplus, L. J. Neuringer, A. Rich, D. C. Steiner and H. Tager for discussion; B. H. Frank and R. E. Chance (Eli Lilly) for biosynthetic human insulin; T. Havel (Harvard Medical School) for the program DG-II and A. Brunger (Yale University) for the program XPLOR; A. Bax, J. Lee, A. G. Redfield, G. Wagner and K. Wüthrich for advice on NMR techniques; and J. Avruch, E. Blout, C. R. Kahn, C. T. Walsh, E. Braunwald and J. T. Potts Jr for encouragement. The coordinates of the DG/SA structures and NMR-derived restraint list will be deposited in the Brookhaven Data Bank.

YY1 is an initiator sequence-binding protein that directs and activates transcription *in vitro*

Edward Seto, Yang Shi* & Thomas Shenk†

Howard Hughes Medical Institute, Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544-1014, USA

REGULATION of eukaryotic messenger RNA transcription is governed by DNA sequence elements that serve as binding sites for sequence-specific transcription factors¹⁻³. These include upstream and downstream promoter-proximal elements, enhancers, repressors, and silencers, which modulate the rate of specific initiation by RNA polymerase II. In addition, the promoter-proximal region between -45 and +30 (relative to the start of initiation) contains two highly conserved motifs, the TATA sequence at around -30 and CA at +1 (ref. 4). Although the TATA element-binding factor TFIID has been purified and cloned from several organisms and has provided invaluable insight into the process of transcription initiation and its regulation, little is known about factors that interact at the +1 region. We have recently shown that the adeno-associated virus type 2 P5 promoter +1 region (P5+1 element) binds transcription factor YY1 (ref. 5). We report here that this sequence is necessary and sufficient for accurate basal transcription. Further, partially purified YY1 can restore basal level transcription from a P5+1 element in a

* Present address: Department of Cellular and Molecular Physiology, Harvard Medical School, Boston, Massachusetts 02115, USA.

† To whom correspondence should be addressed.

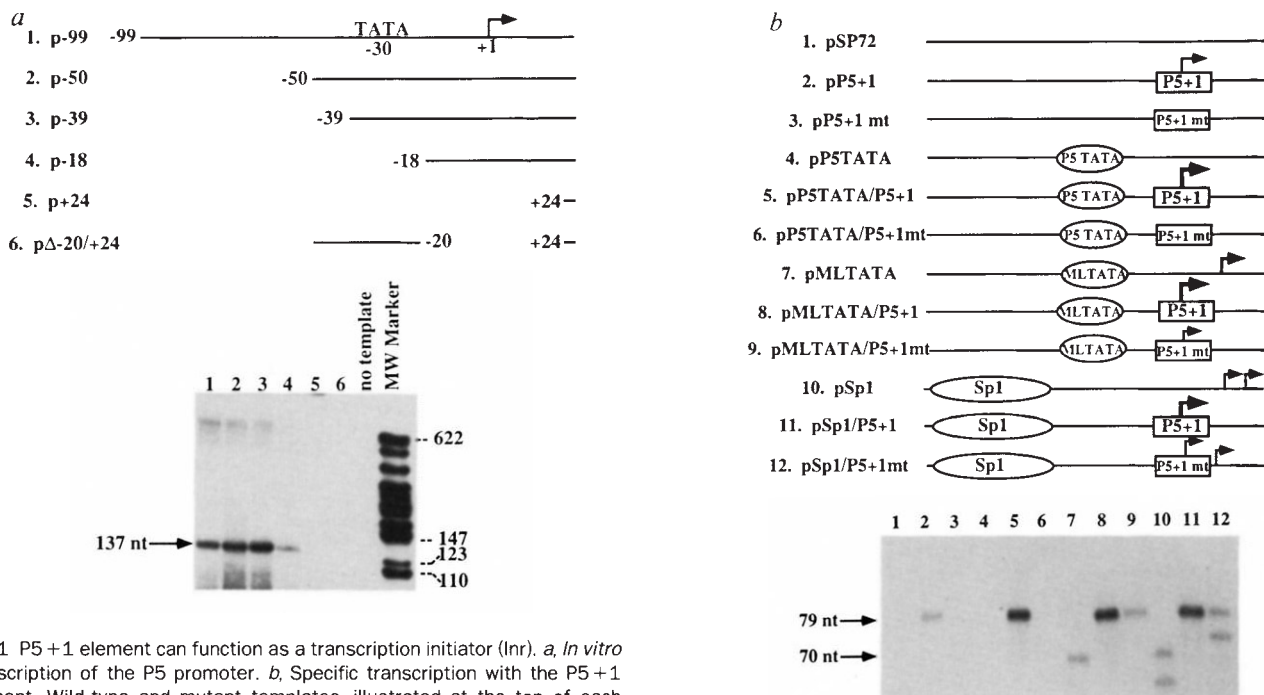


FIG. 1 P5+1 element can function as a transcription initiator (Inr). *a*, *In vitro* transcription of the P5 promoter. *b*, Specific transcription with the P5+1 element. Wild-type and mutant templates, illustrated at the top of each panel, were transcribed in HeLa nuclear extracts, and product RNAs were assayed by reverse transcription. The expected size of the reverse transcripts are shown on the left side of the autoradiogram. Molecular mass markers are derived from ^{32}P -labelled *MspI*-digested pBR322 fragments. METHODS. With the exception of pΔ-20/+24, plasmids in *a* have been described²⁷. Plasmid pΔ-20/+24 was created by deleting the DNA between the *AvaI* and *KpnI* sites in plasmid p-50. Plasmids pMLTATA and pSp1 (referred to as plasmids II and III, respectively, in ref. 7) were provided by Stephen Smale (University of California, Los Angeles). All other plasmids illustrated in *b* were constructed as described for TdT plasmids used in similar studies^{6,7} but with the following oligonucleotides: P5+1, 5'-GATCCCGCTTCAAAATGGAGACCGAGCT-3'; P5+1 mt, 5'-GATCCCGCTTCAAACTTTAGACCGAGCT-3' and their complements to produce *Bam*HI and *SacI* ends; P5TATA, 5'-AATTCTGGGTATTTAAGCCCGAGTGAGGAGCT-3' and its

complement to produce *SacI* and *EcoRI* ends. All constructions were confirmed by DNA sequence analysis²⁸. HeLa nuclear extracts were prepared by the method of Dignam *et al.*²⁹. *In vitro* transcription⁵ and primer extension⁶ were performed as described, except that 400 ng of template DNA was used in each reaction instead. Analysis of RNA in panel *a* used a synthetic oligonucleotide primer complementary to the CAT gene sequence that was present downstream of promoter elements, 5'-GCCATTGGGATATATCAACGGTGG-3'. Primer extension analysis of RNA synthesized in *b* used a synthetic oligonucleotide complementary to the SP6 promoter sequence in the vector (Promega). All experiments were carried out a minimum of three times with at least two separate plasmid and nuclear extract preparations to ensure data reproducibility.

HeLa extract depleted for YY1 or a *Drosophila* embryo extract devoid of YY1 activity, whereas a YY1-specific antibody can block the reactivation. Finally, using electrophoretic mobility shift assay, we have identified YY1-related factors that bind to two other transcription initiators in cellular genes.

We have recently described a human Krüppel-related protein YY1, that interacts with DNA elements centred at -60 and at the transcription initiation region of the adeno-associated virus P5 promoter. As association of YY1 at the +1 region may point to novel means of transcriptional control, we investigated this possibility by examining P5 promoter deletions that define the 5' and 3' boundaries minimally required for accurate and efficient initiation of transcription. Using an *in vitro* transcription assay, DNA templates extending from nucleotide -50 or -39 to +24, both of which include only the TATA and P5+1 elements, were actively transcribed (Fig. 1*a*, lanes 2, 3). An approximately fivefold decrease in transcription was consistently observed when the TATA element was deleted (Fig. 1*a*, lane 4). As expected, no transcription was detected from a promoterless construct (Fig. 1*a*, lane 5). Deletion of sequences from -20 to +24, which removed the P5+1 element but left the TATA box intact, also eliminated detectable transcripts of 113 nucleotides, as expected for RNA products initiated 30 base pairs (bp) downstream from the TATA element (Fig. 1*a*, lane 6). We conclude that the sequence surrounding the P5 transcription start site is sufficient on its own to direct specific initiation of transcription *in vitro*. This minimal sequence is reminiscent of the previously defined terminal deoxynucleotidyltransferase (TdT) gene initiator motif⁶, which contains the transcription

start site. This sequence is sufficient to direct transcription initiation, and its activity is enhanced in the presence of a TATA box.

To confirm that the P5+1 element indeed possesses the ability to function as an initiator, we repeated experiments with the P5+1 sequence similar to those carried out by Smale *et al.*^{6,7}. Oligonucleotides bearing sequences from -7 to +11 in the P5 promoter were inserted in vector pSP72 and examined for the ability to initiate transcription *in vitro*. Plasmid template pP5+1 produced a transcript of the expected length as detected by primer extension analysis (Fig. 1*b*, lane 2). But, plasmid template with a mutation in the P5+1 sequence so that it no longer binds YY1⁵, did not produce detectable specific product (Fig. 1*b*, lane 3). In agreement with previous studies⁷, we found that the adenovirus type 2 major late promoter TATA box or the SV40 Sp1 binding sites alone promoted a low level of transcription initiation (Fig. 1*b*, lanes 7, 10), but the promoter strength and fidelity increased in the presence of an initiator element (Fig. 1*b*, lanes 8, 11). Mutation of the P5+1 element such that it no longer binds YY1 also diminished transcription to its original levels (Fig. 1*b*, lanes 9, 12). In addition, unlike the major late promoter TATA box, we found that the P5 promoter TATA, which is identical to the SV40 early promoter TATA sequence, directed a detectable level of transcription only in the presence of the P5+1 element (Fig. 1*b*, lanes 4, 5).

To determine whether YY1 is required to direct transcription from the P5+1 element, we used DNA affinity chromatography to prepare a HeLa extract depleted of YY1 DNA binding activity (Fig. 2*a*), and asked whether this depleted extract was still capable of transcribing the pP5+1 template. As shown in

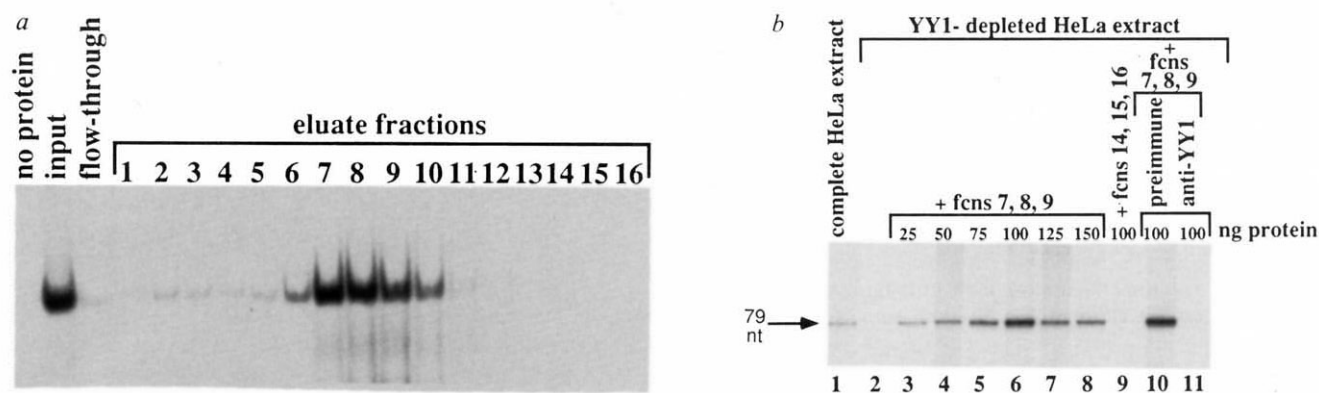
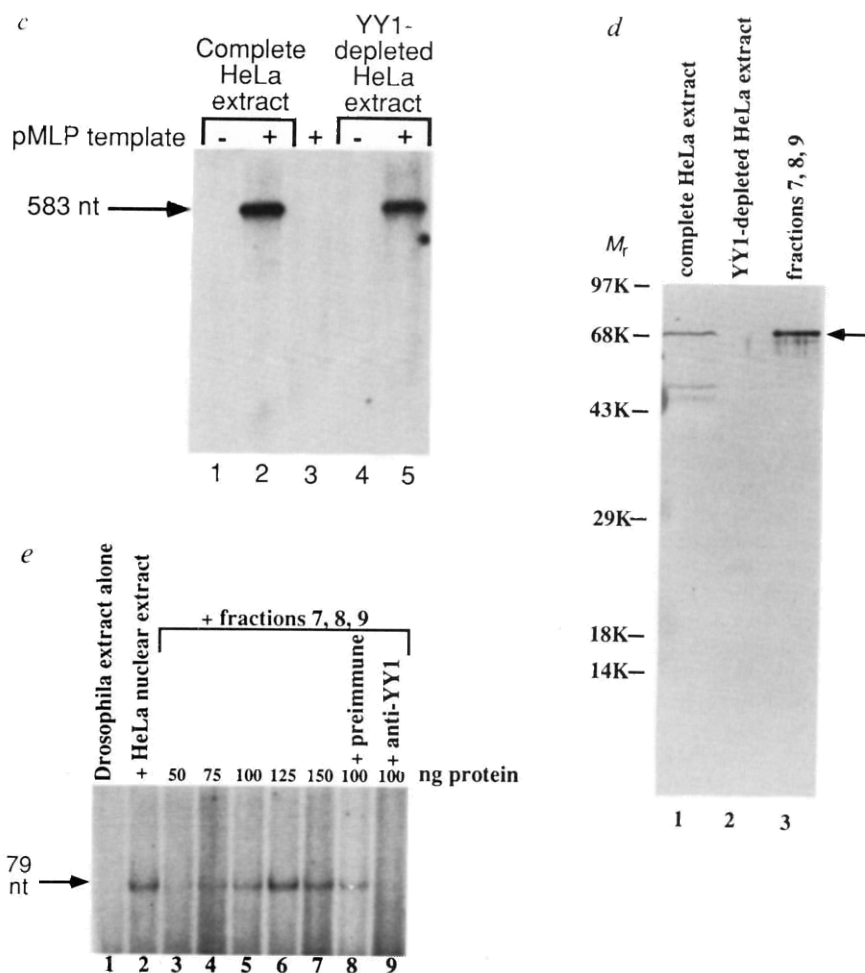


FIG. 2 Transcription directed by the P5 + 1 element in the absence and presence of YY1 activities. *a*, HeLa nuclear extracts were depleted for YY1 activities by two sequential passages through a YY1-specific DNA affinity column. Electrophoretic mobility shift assay (EMSA) was used to monitor YY1 activity. HeLa nuclear extracts depleted for YY1 (*b*, *c*) and *Drosophila* embryo extracts devoid of YY1 (*e*) were used to transcribe template pP5 + 1 (*b*, *e*) or pMLP (*c*). +, reactions performed in the presence, and absence (–) of DNA template. Anti-YY1 or pre-immune antibodies were added where indicated. Primer extension products specific to the template are indicated by arrow. *d*, Western blot analysis of YY1. The arrow indicates the expected size of YY1 protein from HeLa extract. Prestained high-molecular mass markers (BRL) were used as standards.

METHODS. Preparation of DNA-affinity column and purification of YY1 closely followed that method of Kadonaga and Tjian³⁰. Briefly, synthetic oligonucleotides (5'-CGGGAGGGTCTCCATTGTAAG-3' and 5'-CCCTCCCGCTTCAAAATGGAGA-3') were annealed, ligated, then coupled to sepharose columns. Nuclear extracts were prepared from 36 l of HeLa cells grown in spinner culture in SMEM medium supplemented with 5% calf serum and antibiotics. After loading the crude fraction (input), the column was washed and the bound protein eluted as described³⁰. EMSA was used to track YY1 activity. Flow-through from the first passage was reappplied to a fresh column to further deplete YY1 activity. Eluate fractions containing high levels of YY1 activity, as monitored by EMSA (data not shown), were pooled and diluted with 6 volumes of buffer Z (20 mM HEPES pH 7.8, 12.5 mM MgCl₂, 1 mM DTT, 20% glycerol, 0.1% NP40)³⁰, and reappplied to a second column to further purify YY1. Fractions highly enriched in YY1 (panel *a*, eluate fractions 7, 8, 9) from this second pass were pooled, dialysed against buffer D (20 mM HEPES pH 7.9, 20% glycerol, 0.1 M KCl, 0.2 mM EDTA, 0.5 mM PMSF, 0.5 mM DTT)²⁹ and quantified with the Bio-Rad protein assay for use in subsequent *in vitro* transcription and western blot experiments. EMSA using the P5 + 1 oligonucleotides were performed as previously described²⁷ with the exception that the binding reactions contained 12 mM HEPES pH 7.9, 10% glycerol, 5 mM MgCl₂, 60 mM KCl, 1 mM DTT, 50 µg ml⁻¹ BSA, 0.5 mM EDTA, and 0.05% NP40. Eluate fractions were diluted 1:48 before using for EMSA. Plasmid pMLP was constructed by inserting a 797-bp fragment from the adenovirus type 5 major late promoter into the *Sma*/I sites of vector pGEM4 (Promega). For *in vitro* transcription, HeLa nuclear extracts were prepared as described²⁹. *Drosophila* embryo extracts were obtained from Stratagene. Transcription and primer extension were performed as described in Fig. 1b with the following modifications: *In vitro* transcription was performed by preincubating template pP5 + 1 at 0 °C with either buffer D alone (*b*, lane 2 and *e*, lane 1), various amounts of protein fractions (*b*, lanes 3–11; *e*, lanes 3–9), 70 µg of HeLa extract (*e*, lane 2), and 1 µl of either preimmune or



anti-YY1 antibody (*b*, lanes 10, 11; *e*, lanes 8, 9) for 15 min before addition of either HeLa nuclear extract (*b*, lane 1), or YY1-depleted extract (*b*, lanes 2–11), or *Drosophila* embryo extract (*e*, lanes 1–9). Transcription was initiated by addition of ribonucleoside triphosphates and was allowed to proceed for 1 h at 30 °C (*b*) or 30 min at 25 °C (*e*). Transcription reactions presented in *c* contain 100 ng of template DNA. Polyclonal antibody against YY1 was prepared by immunizing rabbits with gel-purified YY1 fusion protein (HIS-YY1) produced in *E. coli* as described⁵. Preimmune and anti-YY1 antibodies were purified using Econo-Pac Protein A columns (Bio-Rad) before use. Proteins (30 µg each, *d*, lanes 1, 2; 0.625 µg, *d*, lane 3) were prepared in SDS denaturation buffer before being separated on a 12.5% SDS-polyacrylamide gel and electroblotted onto a nitrocellulose membrane. The presence of YY1 was determined using the alkaline phosphatase procedure³¹ with anti-YY1 primary antibodies and a biotinylated anti-rabbit secondary antibody.

Fig. 2b, lane 2, transcription from the P5+1 element was barely detectable when the depleted fraction was used as the source of polymerase and transcription factors, although this identical extract supported transcription from an adenovirus 5 major late promoter template (Fig. 2c, lane 5). With the addition of fractions enriched in YY1, a much higher level of transcription was observed (Fig. 2b, lane 3–8). This activation of transcription is highly specific and requires YY1, as fractions with very low YY1 activity were unable to activate transcription (Fig. 2b, lane 9). A YY1-specific antibody was used to ensure that the key factor removed by the DNA affinity chromatography was indeed YY1. The antibody was made by immunization of a rabbit with a YY1 fusion protein produced in *Escherichia coli*, and it can specifically block formation of a YY1-DNA complex in a band shift assay (data not shown). By western blot analysis, this antibody detects a single polypeptide migrating at relative molecular mass of 68,000 (M_r 68K), as expected for YY1, in the fraction retained on the DNA affinity matrix but not in the depleted extract (Fig. 2d, lanes 2, 3). Purified YY1-specific antibody, but not preimmune antibody, was able to reverse the activation effect (Fig. 2b, lanes 10, 11). Taken together, these data strongly suggest that YY1 is an initiator regulatory protein that can activate transcription by binding to the P5+1 element.

Although general RNA polymerase II transcription factors from *Drosophila* and HeLa cells are functionally interchangeable⁸, many sequence-specific DNA-binding proteins that interact with promoter and enhancer elements are not conserved between insects and mammals. YY1 activity is undetectable in *Drosophila* embryo extracts (E.S. and T.S., unpublished results), and, as shown in Fig. 2e, lane 1, *Drosophila* embryo extract alone is unable to support transcription of the P5+1 element. But, when complemented with either a HeLa nuclear extract or with fractions enriched in YY1, transcription was readily observed (Fig. 2e, lane 2–7). Again, transcription required a factor immunologically related to YY1, as YY1-specific antibody blocked the activation (Fig. 2e, lane 9).

In our reconstitution reactions, YY1 produced in *E. coli* cannot replace the semipurified YY1 fractions to stimulate transcription (data not shown). Conceivably, protein modification or a particular polypeptide conformation of YY1 is essential for transcription activation, and YY1 isolated from bacteria may lack these properties. Alternatively, YY1 might require one or more coactivators that serve to bridge to the basal transcription apparatus. Such transcriptional adaptors could copurify with YY1 and be present in the partially purified fractions. Work is now underway to identify these possible coactivator activities.

Cellular factors that bind to the immediate vicinity of the +1 region have previously been identified in several promoters^{9–19}. The mouse mammary tumour virus^{20–22}, adenovirus E1B¹⁰, dihydrofolate reductase¹¹, triosephosphate isomerase¹², and the human immunodeficiency virus type 1^{13,23} promoter+1 regions bear no obvious homology to the P5+1 sequence, and unlike the P5+1 element, these sequences are insufficient for basal transcription in the absence of upstream or downstream elements. The bovine papilloma virus P1 promoter contains a novel Sp1 binding site²⁴ immediately downstream from the start of transcription, and has not yet been tested for its ability to direct transcription independently. The +1 region of the adenovirus major late promoter, by itself, has been variously reported to be able⁶ and unable²⁵ to direct transcription initiation. A DNase footprint spanning the major late promoter TATA box and +1 sequence resulted from the binding of TFIID^{16–18}. More recent data, however, suggest that a novel protein of M_r 85–95K designated the CAP-site binding factor interacts with sequences immediately downstream from the start of transcription in this promoter¹⁹. The ability of this factor to stimulate transcription in the absence of any upstream or downstream sequences remains unknown. By contrast, sequence comparison of the TdT initiator⁶ and the initiation site of the human leukocyte interferon gene (LeIF-J)²⁶ revealed striking similarity to the P5+1

element (Fig. 3a). We therefore examined the ability of the LeIF-J+1 region to function as an initiator and determined whether YY1 binds to either the TdT or the LeIF-J+1 sequences. As shown in Fig. 3b, the LeIF-J+1 sequence functions in initiation of transcription as do those from TdT and the P5.

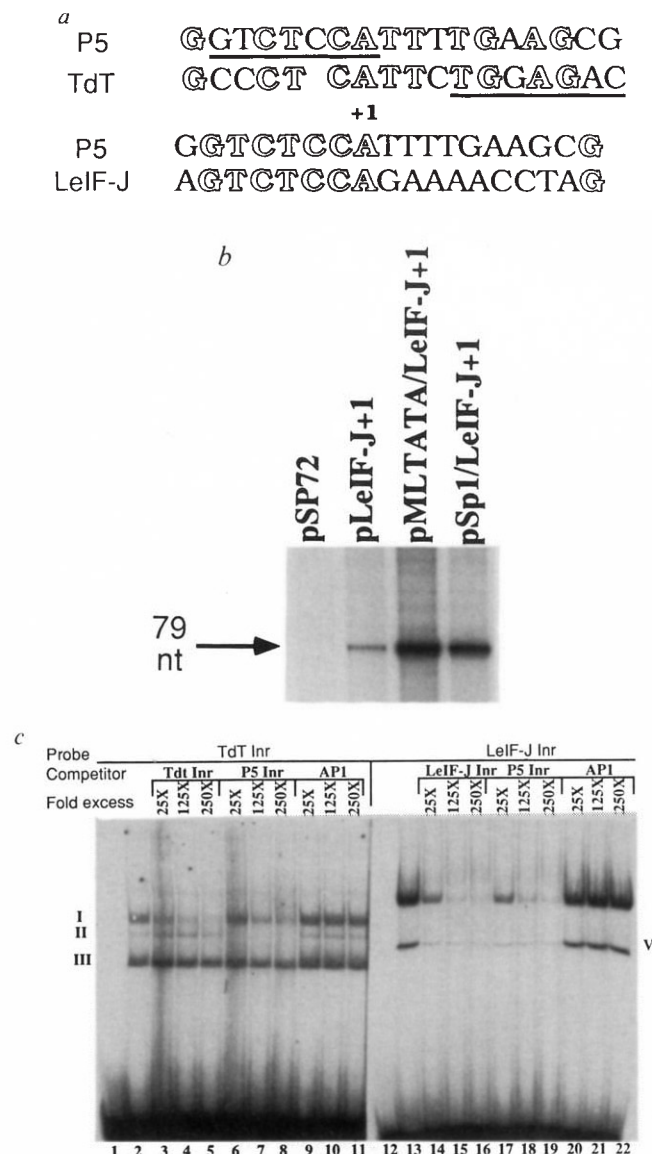


FIG. 3 A YY1-related factor regulates transcription of the TdT and LeIF-J initiators (Inr). **a**, Similarity of initiation sequences from the P5, TdT, and LeIF-J promoters. Open letters indicate sequences conserved between the P5 Inr and the TdT Inr or between the P5 Inr and the LeIF-J Inr. Sequences underlined are inverted repeats. +1 represents the major transcription start site. **b**, *In vitro* transcription of the LeIF-J+1 sequence elements. Reverse transcription products of the RNA are indicated by the arrow. **c**, EMSA of YY1-related factor binding to the TdT and LeIF-J Inr. Formation of complex I is specifically inhibited by addition of excess TdT Inr and the P5 Inr oligonucleotides but not by the addition of an AP1 oligonucleotide. Complex II and III are nonspecific. Complex IV is specifically competed by addition of excess LeIF-J Inr and the P5 Inr oligonucleotides but not by the addition of an AP1 oligonucleotide. Complex V is nonspecific.

METHODS. Construction of plasmids for *in vitro* transcription were similar to those described in Fig. 1b. Oligonucleotides used were: LeIF-J+1, 5'-GATCCCTAGGTTTCTCGAGACTGAGCT-3' and its complement to produce *Bam*HI and *Sac*I ends. EMSA was carried out as in Fig. 2a, and probes used for the assay were ³²P-5' end-labelled double-stranded oligonucleotides whose sequences are given in **a**. Molar excess of unlabelled competitor was added as indicated. Nonspecific AP1 competitor consisted of oligonucleotides 5'-GGATGTTATAAAGCATGAGTCAGACACCTCTGGCT-3' and its complement. No proteins were added in reactions presented in **c**, lanes 1, 12.

Moreover, an oligonucleotide consisting of the P5+1 sequence can effectively inhibit the binding of a nuclear factor to the TdT (Fig. 3c, lanes 6–8, complex I) as well as the LeIF-J Inr sequences (Fig. 3c, lanes 17–19, complex IV). We suspect that there may be multiple types of initiator elements, with those of P5, TdT, and the LeIF-J belonging to the same class. Evidence for this comes from the facts that (1) a variety of electrophoretic mobility shift complexes were observed on comparison of a series of initiator sequences (data not shown), (2) the complexes generated using binding sites corresponding to P5, TdT, and LeIF-J did not migrate identically, and (3) antisera directed against YY1 had no effects on complexes generated by TdT and LeIF-J Inr, even though they were each able to compete for YY1 binding (Fig. 3c; data not shown).

This is the first identification of a factor that can mediate transcription through an initiator element. YY1 can direct the general transcription machinery to initiate RNA synthesis at its binding site. Previously, we have shown that the P5+1 element when placed upstream of either a synthetic promoter or the SV40 early promoter/enhancer can repress transcription⁵. Here, we show that the same element can activate transcription when present alone or downstream from the TATA or Sp1 sites. This activation/repression phenomenon is not unprecedented: Roeder and coworkers¹⁴ recently described an activity (LBP-1) that repressed transcription when bound to an upstream site overlapping the TATA element in the HIV-1 promoter, but enhanced transcription when bound at the +1 region. Kato *et al.*¹⁴ suggested that this activity may cooperate with other factors at the +1 site to enhance transcription but binding of LBP-1 to upstream sites blocked binding of TFIID to the TATA element, thereby suppressing transcription in a dominant negative fashion. Whether YY1 works in a similar manner remains to be determined.

Note added in proof: While this paper was under review we learned that three additional groups have cloned YY1. N. Hariharan, D. E. Kelley and R. P. Perry³² have termed the protein δ , and demonstrated that it binds to sequence elements downstream of the transcriptional start sites in ribosomal protein genes where it stimulates transcription. K. Park and M. L. Atchison³³ termed it NF-E1, and showed that it binds to the IgK 3' enhancer at a site that negatively regulates transcription. J. R. Flanagan *et al.*³⁴ termed it UCRBP, and determined that it binds in the long terminal repeat of Moloney leukaemia virus where it inhibits transcription. □

31. Harlow, E. & Lane, D. *Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1988).
32. Hariharan, N., Kelley, D. E. & Perry, R. P. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
33. Park, K. & Atchison, M. L. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
34. Flanagan, J. R. *et al.* *Proc. natn. Acad. Sci. U.S.A.* (in the press).

ACKNOWLEDGEMENTS. We thank S. Smale for plasmids and C. Parks and U. Müller for advice, discussion, and critical reading of the manuscript. This work was supported by a grant from the National Cancer Institute. E.S. and T.S. are American Cancer Society postdoctoral fellow and research professor, respectively.

Cooperative interaction of an initiator-binding transcription initiation factor and the helix-loop-helix activator USF

Ananda L. Roy, Michael Meisterernst, Philippe Pognonec & Robert G. Roeder

Laboratory of Biochemistry and Molecular Biology,
The Rockefeller University, New York, New York 10021, USA

TRANSCRIPTION initiation by mammalian RNA polymerase II is effected by multiple common factors^{1,2} interacting through minimal promoter elements and regulated by gene-specific factors³ interacting with distal control elements. Minimal promoter elements that can function independently or together, depending on the specific promoter, include the upstream TATA box^{4,5} and a pyrimidine-rich initiator^{6–8} (Inr) overlapping the transcription start site. The binding of TFIID to the TATA element^{4,9} promotes the assembly of other factors into a preinitiation complex^{10–12} but factors which function at the Inr have not been defined. We show here that a novel factor (TFII-I) binds specifically to Inr elements, supports basal transcription from the adenovirus major late promoter and is immunologically related to the helix-loop-helix activator USF (ref. 13). We further show that TFII-I also binds to the upstream high-affinity USF site (E box), that USF also binds to the Inr, and that TFII-I and USF interact cooperatively at both Inr and E box sites. Thus, TFII-I represents a novel type of transcription initiation factor whose interactions at multiple promoter elements may aid novel communication mechanisms between upstream regulatory factors and the general transcriptional machinery.

Inr and Inr-like elements in adenovirus major late (AdML), human immunodeficiency virus-1 (HIV-1) and terminal deoxynucleotidyl transferase (TdT) promoters are summarized in Fig. 1a. Electrophoretic mobility shift assays with an AdML Inr 1 (ML-I) probe revealed several HeLa cell components that bound specifically (data not shown). On further purification one Inr-binding component, designated TFII-I, was identified as a polypeptide of relative molecular mass 120,000 (M_r 120K) (Fig. 1b, c). A second weakly binding component was identified as USF, a cellular factor which activates the AdML promoter through a high-affinity E box sequence (CACGTG) at around –60 (refs 9, 14, 15); this was verified by analysis of the recombinant 43K USF polypeptide (USF⁴³) (Fig. 1d). Competition studies with oligonucleotides containing wild type versus mutant Inr elements, as well as DNase I footprint analyses (Fig. 3b, data not shown), demonstrated specific binding of both TFII-I (Fig. 1c) and USF⁴³ (Fig. 1d) not only to the AdML Inr elements, but also to HIV-1 and TdT Inr elements. Consistent with the similar DNA binding specificities of TFII-I and USF⁴³ (see below), both proteins reacted specifically with an anti-USF⁴³ polyclonal antibody (Fig. 1b). This suggests that these proteins may be structurally related, and possibly in the same family.

Having established the Inr-binding properties of TFII-I, we investigated its potential for transcriptional activation of the

Received 25 June; accepted 24 September.

1. Dynan, W. S. & Tjian, R. *Nature* **316**, 774–778 (1985).
2. Ptashne, M. *Nature* **335**, 683–689 (1988).
3. Mitchell, P. J. & Tjian, R. *Science* **245**, 371–378 (1989).
4. Bucher, P. & Trifonov, E. *N. Nucleic Acids Res.* **14**, 10009–10026 (1986).
5. Shi, Y., Seto, E., Chang, L.-S. & Shenk, T. *Cell* **67**, 377–388 (1991).
6. Smale, S. T. & Baltimore, D. *Cell* **57**, 103–113 (1989).
7. Smale, S. T., Schmidt, M. C., Berk, A. J. & Baltimore, D. *Proc. natn. Acad. Sci. U.S.A.* **87**, 4509–4513 (1990).
8. Heiermann, R. & Pongs, O. *Nucleic Acids Res.* **13**, 2709–2730 (1985).
9. Cordingley, M. G., Riegel, A. T. & Hager, G. L. *Cell* **48**, 261–270 (1987).
10. Parks, C. L., Banerjee, S. & Spector, D. J. *J. Virol.* **62**, 54–67 (1988).
11. Means, A. L. & Farnham, P. J. *Molec. cell. Biol.* **10**, 653–661 (1990).
12. Boyer, T. G. & Maquat, L. E. *J. Biol. Chem.* **265**, 20524–20532 (1990).
13. Jones, K. A., Luciw, P. A. & Duchange, N. *Genes Dev.* **2**, 1101–1114 (1988).
14. Kato, H., Horikoshi, M. & Roeder, R. G. *Science* **251**, 1476–1479 (1991).
15. Stenlund, A., Bream, G. L. & Botchan, M. R. *Science* **236**, 1666–1671 (1987).
16. Sawadogo, M. & Roeder, R. G. *Cell* **43**, 165–175 (1985).
17. Van Dyke, M. W., Roeder, R. G. & Sawadogo, M. *Science* **241**, 1335–1338 (1988).
18. Nakajima, N., Horikoshi, M. & Roeder, R. G. *Molec. cell. Biol.* **8**, 4028–4040 (1988).
19. Garfinkel, S., Thompson, J. A., Jacob, W. F., Cohen, R. & Safer, B. *J. Biol. Chem.* **265**, 10309–10319 (1990).
20. Buetti, E. & Diggelmann, H. *EMBO J.* **2**, 1423–1429 (1983).
21. Majors, J. & Varmus, H. E. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5866–5870 (1983).
22. Lee, et al. *Nucleic Acids Res.* **12**, 4191–4206 (1984).
23. Okamoto, T., Benter, T., Josephs, S. F., Sadaie, M. R. & Wong-Staal, F. *Virology* **177**, 606–614 (1990).
24. Li, R., Knight, J. D., Jackson, S. P., Tjian, R. & Botchan, M. R. *Cell* **65**, 493–505 (1991).
25. Hen, R., Sassone-Corsi, P., Corden, J., Gaub, M. P. & Chambon, P. *Proc. natn. Acad. Sci. U.S.A.* **79**, 7132–7136 (1982).
26. Ullrich, A., Gray, A., Goeddel, D. V. & Dull, T. J. *J. molec. Biol.* **156**, 467–486 (1982).
27. Chang, L., Shi, Y. & Shenk, T. *J. Virol.* **63**, 3479–3488 (1989).
28. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
29. Dignam, J. D., Lebovitz, R. M. & Roeder, R. G. *Nucleic Acids Res.* **11**, 1475–1489 (1983).
30. Kadonaga, J. T. & Tjian, R. *Proc. natn. Acad. Sci. U.S.A.* **83**, 5889–5893 (1986).

AdML core promoter in a system reconstituted with RNA polymerase II and the general factors TFIIA, TFIIB, TFIID³⁸ and TFIIE/F (ref. 16). Basal transcription was undetectable in a system reconstituted with all these factors except TFIIA but could be stimulated by purified TFII-I (Fig. 2a, lanes 1–5) or by TFIIA (ref. 16; data not shown). Moreover, the selective inhibition of TFII-I-dependent transcription by an oligonucleotide with an intact Inr site (Fig. 2b, lanes 1–7) demonstrated that the binding of TFII-I is necessary for core promoter activation. These results clearly establish a transcriptional function for TFII-I and indicate that template interactions are important for this function. They also indicate a pathway for preinitiation complex assembly distinct from that which uses TFIIA, which is structurally distinct from TFII-I (J. DeJong, A.L.R. and R.G.R., unpublished observations) and interacts with the preinitiation complex in a sequence-independent manner^{11,12}.

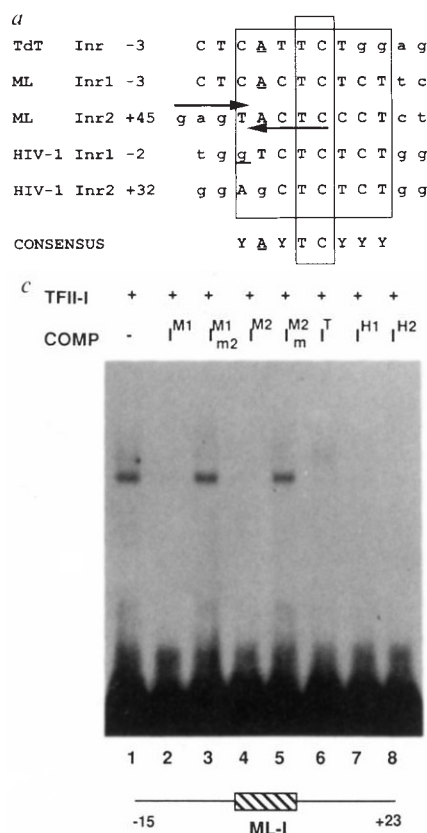
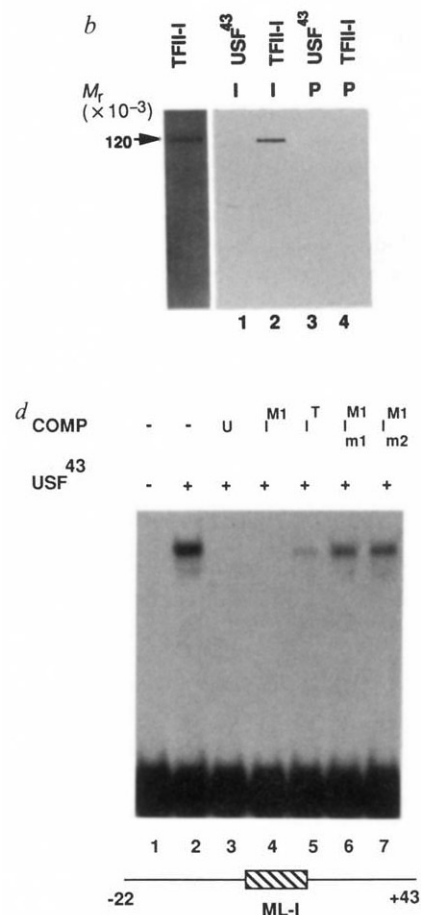


FIG. 1 Size, antigenic relatedness and initiator-binding properties of TFII-I and USF⁴³. *a*, Nucleotide sequence of Inr elements. Capital letters, bases which match the consensus, except for TdT where they indicate the Inr defined by mutation⁸. The initiating bases are underlined. An eight-base pair palindromic sequence in ML Inr 2 is indicated by arrows. TFII-I binding to each of these elements was shown by electrophoretic mobility shift assay (EMSA) (as in *c*) and, in some cases, by DNase I footprint as well. *b*, SDS-PAGE analysis (left) showing a single (silver-stained) polypeptide of 120K in HPLC-purified TFII-I and western blot analysis (right) showing crossreactivity of the 120K polypeptide (and USF⁴³) with anti-USF⁴³ immune (I) but not preimmune (P) serum. The 120K polypeptide coeluted with Inr-binding and Inr-dependent transcription activities during gradient elution in the final HPLC purification step. The size of the silver-stained and immunoreactive polypeptides were estimated from coelectrophoresed markers (not shown). Blockage of the anti-USF⁴³ antiserum with homogeneous USF⁴³ eliminated crossreactivity with both p120 and USF⁴³, further demonstrating the specificity of the reactivity (data not shown). *c*, EMSA showing specific binding of TFII-I to AdML, HIV-1 and TdT Inr elements. Reactions were done with HPLC-purified TFII-I, the indicated AdML DNA probe containing Inr 1, and unlabelled oligonucleotide competitors (comp) I^{M1} (ML Inr 1, sequences -12 to +18), I^{M2} (I^{M1} with mutations T to A at +3, C to T at +4, and T to A at +5), I^{M3} (ML Inr 2, sequences +24 to +67), I^{M4} (I^{M2} with mutations G to C at +42, G to C at +46, A to G at +48, C to G at +53, T to A at +57, and G to C at +58), I^T (TdT Inr, sequences -12 to +18), I^{H1} (HIV-1 Inr 1, sequences -12 to

Given that TFII-I and USF bind independently to the Inr site, as well as indications of structural relatedness, we investigated the possibility that they might interact cooperatively at this site. Mobility shift assays with a longer ML-I probe generated two TFII-I-dependent complexes that required an intact Inr (Fig. 3a) and appeared from other studies to reflect monomeric (C2) and dimeric (C1) species (data not shown). At the low concentration used, USF⁴³ alone showed barely detectable binding. But the simultaneous addition of USF⁴³ and TFII-I generated a novel complex (Fig. 3a) indicative of highly cooperative interactions. A higher-resolution analysis (Fig. 3a) showed



+23), I^{H2} (HIV-1 Inr 2 sequences +23 to +60). *d*, EMSA showing specific binding of USF⁴³ to AdML and TdT Inr elements. Reactions contained purified TFII-I, the indicated AdML DNA probe containing Inr 1, and unlabelled oligonucleotide competitors I^{M1} (as in *c*), I^{M2} (as in *c*), I^{M3} (I^{M1} with mutations C to G at -2, T to A at -3, and C to G at -4), and U (ML-U sequences -75 to -35). Similar competition studies, as well as direct analyses with labelled probes, also showed specific binding of USF⁴³ to the AdML Inr 2 and the HIV-1 Inr 1 and Inr 2 elements (data not shown).

METHODS. TFII-I was purified to apparent homogeneity from HeLa nuclear extracts by standard ion exchange, DNA affinity, and HPLC methods to be published elsewhere. USF⁴³ was purified to 95% homogeneity by ammonium sulphate fractionation and HPLC chromatography on DEAE-5PW (BioRad)²³, after expression of the corresponding complementary DNA in bacteria using the Studier T7 vector system. The western blot analysis was done according to standard protocols using 1:1,000 dilutions of antibody. EMSAs were done as follows: the protein was incubated for 20 min at 30 °C with 0.1–0.2 ng labelled probe (adjusted to 1×10^4 c.p.m. ng⁻¹) in a reaction mixture (20 μ l) containing 20 mM Tris, pH 7.9, 10% (v/v) glycerol, 10 mM EDTA, 5 mM DTT, 80 mM KCl, 200 ng poly dA:poly dT as nonspecific DNA and BSA to bring the total protein concentration to 500 ng. Equimolar amounts of oligonucleotide competitors (50-fold molar excess over probe) were added as indicated. The reaction products were analysed by electrophoresis in 4% native polyacrylamide gels.

TABLE 1 Relative stabilities of complexes of TFII-I and USF⁴³ at various sites

Factor	Site	K_B (10^{-9})
TFII-I	ML-U	4.90
TFII-I	ML-I1	3.10
TFII-I	ML-I2	5.80
USF	ML-U	2.40
USF	ML-I1	0.12
USF	ML-I2	0.20

Binding constant values (K_B) were determined by standard filter binding assay, using oligonucleotide probes of equivalent lengths and specific activity. These contained ML sequences -75 to -35 (ML-U), -15 to +23 (ML-I1) and +24 to +65 (ML-I2).

that the complex formed by the cooperative USF/TFII-I interactions migrates between the two TFII-I complexes and slightly slower than the USF⁴³ complex. Consistent with the involvement of both USF and TFII-I in the novel complex, it was eliminated or markedly reduced by oligonucleotides containing, respectively, the high-affinity USF site (ML-U) or the Inr 1 element (ML-I). The ML-U oligonucleotide did not interfere with the USF-independent binding of TFII-I (Fig. 3a), in agreement with the inability of this oligonucleotide to inhibit TFII-I function (Fig. 2). Cooperative interactions between USF⁴³ and TFII-I were confirmed by DNase I footprint analysis which revealed patterns (areas of protection and hypersensitive sites) quantitatively and qualitatively distinct from those observed with either factor alone (Fig. 3b, data not shown). These results clearly indicate cooperative interactions between TFII-I and USF⁴³ at the ML-I site and are most consistent with formation of a heteromeric complex containing both components, especially in light of the distinct mobility of the complex and the 20-fold lower affinity of USF alone, relative to TFII-I, for this site (Table 1).

Subsequent mobility shift and oligonucleotide competition analyses showed that TFII-I also binds independently and specifically to the high-affinity USF binding site (ML-U) of the ML promoter (Fig. 3c). Moreover, a combination of TFII-I and USF⁴³ resulted in highly cooperative interactions at this site, enhancing the formation of a complex with a mobility similar to that of the USF⁴³ complex. The resulting complex may reflect only a TFII-I-dependent stimulation of USF⁴³ binding. But other observations are more consistent with the possibility that the enhanced complex is a heteromeric USF⁴³/TFII-I complex that fortuitously migrates with the USF⁴³ complex. First, the increased amount of the new complex is accompanied by a decreased amount of the TFII-I complex under conditions of probe excess. Second, like the TFII-I complex (Fig. 3c), but unlike the USF⁴³ complex (data not shown, compare with Fig. 1d), this enhanced complex is much more resistant to a competing ML-I oligonucleotide than to a competing ML-U oligonucleotide.

The novel observations include: a 120K USF-related factor, TFII-I, which shows Inr-dependent promoter binding and transcriptional activation; independent interactions of TFII-I with the upstream high-affinity USF site and of USF with the Inr site; and cooperative interactions between USF and TFII-I at both the Inr and the high-affinity USF site. These observations provide direct evidence for a novel transcription initiation factor and corresponding promoter-activation pathway and they suggest novel mechanisms for communication between the basic transcriptional machinery and at least one important class of regulatory proteins (helix-loop-helix proteins).

Although quite important from the standpoint of indicating a new promoter-activation pathway that might be selectively targeted by some regulatory factors, the identification of a novel initiation factor interacting with the Inr was not unexpected. Much more surprising were the interaction of TFII-I at the

high-affinity USF site (E box), the interaction of USF at the Inr, and the cooperative interactions of both factors at both sites. Direct functional roles for TFII-I interactions at E box sites and for USF interactions at Inr elements have not yet been demonstrated *in vitro*, although transfection studies have shown that ectopic USF⁴³ can stimulate transcription of reporter genes through Inr elements (H. Du, A. L. R. and R. G. R., unpublished observations). The ability of USF⁴³ and TFII-I to bind two distinct sequence elements (affinities summarized in Table 1) is also quite intriguing, but not yet understood from a structural standpoint. But one feature of the TFII-I interactions is that binding to the Inr was refractory to competition by ML-U sequences whereas binding to the ML-U site was relatively resistant to competition by ML-I sequences. These observations raise the possibility that TFII-I binds to the ML-U and ML-I sites through distinct domains on the protein, and that binding of one domain to its cognate site is not inhibited by the presence of an oligonucleotide that binds to the second domain. The presence of distinct DNA-binding domains would expand the potential for simultaneous interactions of TFII-I at distinct promoter elements.

The aforementioned cooperative interactions of TFII-I and USF are especially interesting with regard to their potential importance for regulation. USF⁴³ is a member of the Myc family of helix-loop-helix (HLH) proteins^{13,17,18}. This fact, the immunological relationship and physical interactions between USF⁴³ and TFII-I, and the interaction of both proteins with an E motif (ML-U site) typically recognized by many HLH regulatory proteins¹⁷⁻²¹ raises the possibility that TFII-I might also be an HLH protein. If so, many HLH regulatory proteins might

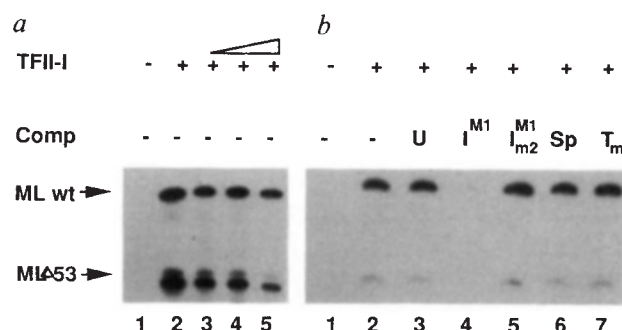


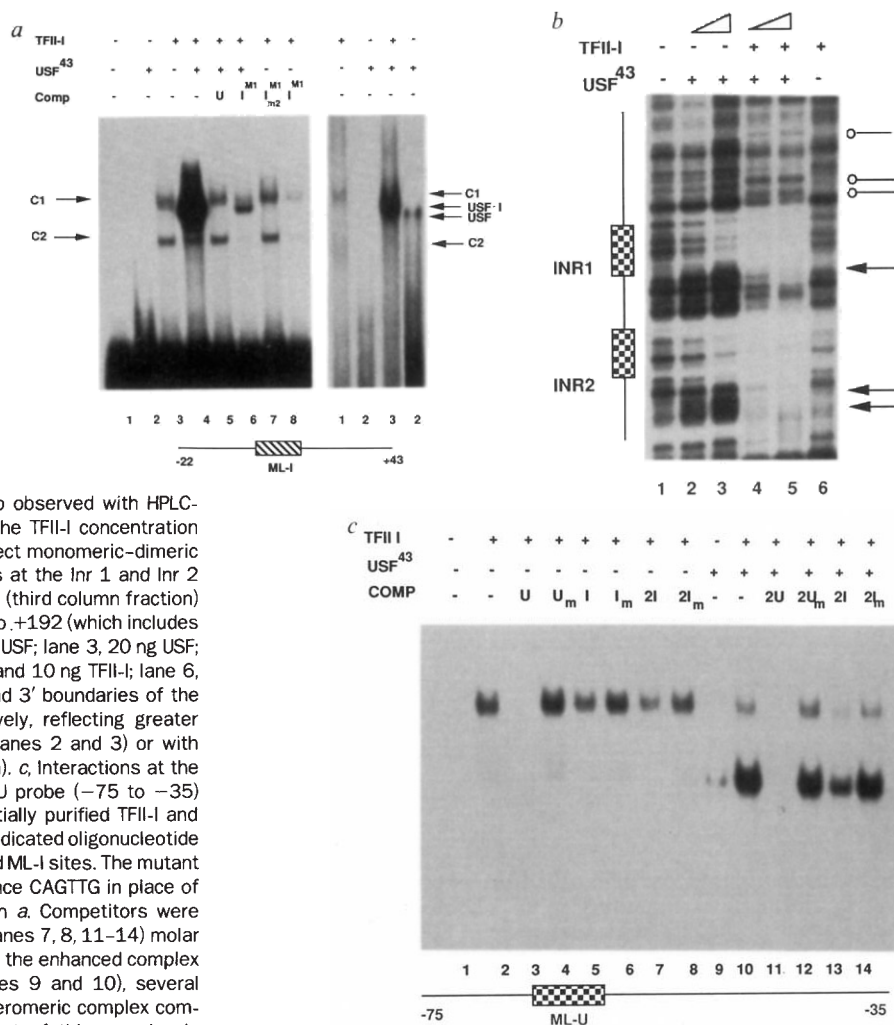
FIG. 2 TFII-I-dependent transcription from the AdML core promoter. All reactions contained common factors TFIIIB, recombinant TFIIID³⁸, TFIIIE/F, RNA polymerase II, and AdML wild type and Δ53 templates. The Δ53 template contains only core promoter elements (TATA box and Inr 1 site) whereas the wild-type template contains the upstream USF site as well. But the use of the recombinant 38K TATA-binding polypeptide (TFIIID³⁸) in place of natural TFIIID (ref. 24) and the absence of other cofactors¹⁶ precludes activation through the upstream USF sites. *a*, ML promoter activation by purified TFII-I. Transcription reactions contained no additional factors (lane 1), a partially purified TFII-I (lane 2), or highly purified TFII-I (lanes 3–5) in increasing amounts (1, 2 and 4 ng, respectively). *b*, Selective inhibition of TFII-I-dependent transcription by Inr-containing oligonucleotides. Transcription reaction contained no additional factors (lane 1) or partially purified TFII-I and equivalent amounts (10 ng) of oligonucleotide competitors containing the ML-U site (U, lane 3), the ML Inr 1 site (I^{M1}, lane 4), a mutated ML Inr 1 site (I_{m2}^{M1}, lane 5), and SP1 site (SP, lane 6) or a mutated TATA element (T_m, lane 7). The fact that the Inr-containing oligonucleotide selectively inhibits both core promoter transcription (*b*, lane 4) and TFII-I binding to Inr elements (Fig. 1c, lane 2) indicates that these activities reside in the same protein and that binding to the promoter Inr is essential for function.

METHODS. Reactions for *in vitro* transcription were as described¹⁶ and contained, in order of addition: 50 ng of template; TFII-I as indicated; a mixture of TFIIIB, recombinant TFIIID³⁸ (cloned TATA-binding polypeptide of human TFIIID), TFIIIE/F and RNA polymerase II in the amounts described previously¹⁶. After incubation at 30 °C for 1 h, reactions were processed and labelled transcripts analysed on 5% urea gels. AdML wild type (5' sequences to -400) and Δ53 (5' sequences to -53 and lacking the ML-U site) promoter sequences were attached to G-less cassettes^{9,16}.

FIG. 3 TFII-I and USF⁴³ bind independently and interact cooperatively at both the AdML Inr and upstream USF (E box) sites. Cooperativity is defined as an increase in the number of promoter-bound complexes with both factors present in comparison to the number observed with either factor alone. *a*, Interactions at the Inr 1 site. EMSAs in the left panel were done with an AdML probe (−22 to +43) containing Inr 1, USF⁴³ (2 ng) and/or partially purified TFII-I, and oligonucleotide competitors as indicated. Competitors containing the ML-U (U) or either wild type (I^{M1}) or mutant (I^{M2}) ML Inr 1 sites were as described in Fig. 1 and were present in 50-fold molar excess over the probe. EMSAs in the right panel were done similarly except that the amount of TFII-I was lowered two-fold and the gel was run longer. Lane 2* is a longer exposure of lane 2 to more clearly visualize the USF⁴³ complex and to demonstrate that it is distinct (higher mobility) from the USF-TFII-I complex.

Results identical to those described here were also observed with HPLC-purified TFII-I. The ratio of C2 to C1 increased as the TFII-I concentration was decreased, indicating that these complexes reflect monomeric-dimeric (or higher order) interactions of TFII-I. *b*, Interactions at the Inr 1 and Inr 2 sites. DNase I footprint analysis with USF⁴³ and TFII-I (third column fraction) on the AdML probe containing sequences from −53 to +192 (which includes both Inr 1 and 2). Lane 1, no additions; lane 2, 10 ng USF; lane 3, 20 ng USF; lane 4, 10 ng USF and 5 ng TFII-I; lane 5, 10 ng USF and 10 ng TFII-I; lane 6, 5 ng TFII-I. In the presence of both factors the 5' and 3' boundaries of the protected regions lie near −3 and +70, respectively, reflecting greater protection than observed with either USF⁴³ alone (lanes 2 and 3) or with higher concentrations of TFII-I alone (data not shown). *c*, Interactions at the upstream E box site. EMSA was done with an ML-U probe (−75 to −35) containing the high-affinity USF site (CACGTG)⁹. Partially purified TFII-I and USF⁴³ (1 ng) were added as indicated, along with the indicated oligonucleotide competitors containing wild type and mutant ML-U and ML-I sites. The mutant ML-U oligonucleotide (U_m) contained the core sequence CAGTTG in place of CACGTG, whereas other oligonucleotides were as in *a*. Competitors were present in either a 50-fold (lanes 3–6) or 100-fold (lanes 7, 8, 11–14) molar excess relative to the probe. Although the mobility of the enhanced complex is similar to that of the USF⁴³/ML-U complex (lanes 9 and 10), several observations suggest that it represents a novel heteromeric complex comprising USF⁴³ and TFII-I. First, the increased amount of this complex is accompanied by a corresponding decrease in the amount of the TFII-I/ML-U complex in probe excess. Second, in contrast to the USF⁴³/ML-U complex (data not shown, compare with Fig. 1*d*), both the enhanced complex and the clearly distinguished TFII-I/ML-U complex are much more resistant to competing ML-I sequences than to competing ML-U sequences (Fig. 3*c*).

METHODS. EMSA conditions were as in Fig. 1. Footprinting reactions were



also done under essentially the same conditions, except that the DNA probe (−53 to +192) was prepared by 3' end-labelling (upper strand) with [³²P]ATP and Klenow enzyme. After the initial incubation, 15 ng DNase I was added and the reactions were incubated for an additional 1 min at room temperature, before processing and analysis. The TFII-I used was purified through three chromatographic steps.

interact with TFII-I (or possibly USF⁴³) through classical E box motifs or through Inr-like elements, including the oncoprotein Myc or its known partners, as the only known recognition site for these proteins is in essence the ML-U site^{17,18}. If TFII-I were to have a general role in modulating the HLH family of regulators, the positive or negative effects of these proteins might be targeted directly toward transcription initiation and the general factors (through TFII-I).

The distribution and possible diversity of Inr-like elements

among various promoters is not yet clear. But although a strict sequence conservation is missing in many of the known Inr elements, sequence similarities to the classical TdT initiator^{7,8} are apparent and suggest the degenerate consensus YAYTCYYY (Fig. 1*a*). The positions of the pyrimidine-rich Inr-like elements are not necessarily fixed and may be found distal to the initiation site²², which may also indicate context effects and the possibility for interactions of Inr-bound factors with a broader variety of control factors.

Received 10 July; accepted 7 October 1991.

- Matsui, T., Segall, J., Weil, P. A. & Roeder, R. G. *J. Biol. Chem.* **255**, 11992–11996 (1980).
- Sumimoto, H., Ohkuma, Y., Yamamoto, T., Horikoshi, M. & Roeder, R. G. *Proc. natn. Acad. Sci. U.S.A.* **87**, 9158–9162 (1990).
- Maniatis, T., Goodbourn, S. & Fischer, J. A. *Science* **236**, 1237–1245 (1987).
- Nakajima, N., Horikoshi, M. & Roeder, R. G. *Molec. cell. Biol.* **8**, 4028–4040 (1988).
- Breathnach, R. & Chambon, P. A. *Rev. Biochem.* **50**, 349–383 (1981).
- Grosschedl, R. & Birnstiel, M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1432–1436 (1980).
- Smale, S. T. & Baltimore, D. *Cell* **57**, 103–113 (1989).
- Smale, S. T., Schmidt, M. C., Berk, A. J. & Baltimore, D. *Proc. natn. Acad. Sci. U.S.A.* **87**, 4509–4513 (1990).
- Sawadogo, M. & Roeder, R. G. *Cell* **43**, 165–175 (1985).
- Van Dyke, M., Roeder, R. G. & Sawadogo, M. *Science* **241**, 1335–1338 (1988).
- Buratowski, S., Hahn, S., Guarente, L. & Sharp, P. A. *Cell* **56**, 549–561 (1989).
- Maldonado, E., Ha, I., Cortes, P., Weis, L. & Reinberg, D. *Molec. cell. Biol.* **10**, 6335–6347 (1990).
- Gregor, P. D., Sawadogo, M. & Roeder, R. G. *Genes Dev.* **4**, 1730–1740 (1990).
- Carthew, R. W., Chodosh, L. A. & Sharp, P. A. *Cell* **43**, 439–448 (1985).

- Miyamoto, N. G., Moncollin, V., Egly, J. M. & Chambon, P. *EMBO J.* **4**, 3563–3570 (1985).
- Meisterernst, M., Roy, A. L., Lieu, H. L. & Roeder, R. G. *Cell* **66**, 1–20 (1991).
- Blackwood, E. M. & Eisenman, R. N. *Science* **251**, 1149–1280 (1991).
- Prendergast, G. C., Lawe, D. & Ziff, E. B. *Cell* **65**, 395–407 (1991).
- Murre, C., Schonleber-McCaw, P. & Baltimore, D. *Cell* **56**, 777–783 (1989).
- Murre, C. *et al.* *Cell* **58**, 537–544 (1989).
- Davis, R. L., Cheng, P. F., Lassar, A. B. & Weintraub, H. *Cell* **60**, 733–746 (1990).
- Nakatani, Y. *et al.* *Nature* **348**, 86–88 (1990).
- Pognonec, P. & Roeder, R. G. *Molec. cell. Biol.* **11**, 5125–5136 (1991).
- Hoffman, A. *et al.* *Nature* **346**, 387–390 (1990).

ACKNOWLEDGEMENTS. We thank M. Horikoshi for the gift of cloned human TFII-I, H. Sumimoto for help with HPLC fractionation, B. Moorefield for critically reading the manuscript. A.L.R. is a Damon Runyon-Walter Winchell Cancer Research Fund postdoctoral fellow. M.M. was supported by a fellowship from Deutsche Forschungsgemeinschaft. P.P. was supported by an EMBO fellowship. This study was supported by NIH grants (to R.G.R.) and by general support from the Pew Trusts to the Rockefeller University.

ZEKE threshold photoelectron spectroscopy

Edward R. Grant and Michael G. White

New methods for the precise (state-detailed) resolution of ionization thresholds open new vistas in photoionization dynamics, ion spectroscopy and state-selected ion-molecule reactions.

CHEMISTS and molecular physicists from Europe, North America and Japan met last month at a European Research Conference in Kreuth, near Munich, to discuss recent advances in the selectivity and resolution of methods for probing the spectroscopy and photoionization dynamics of small molecules. An important focus early in the development of chemical physics, photoionization has become a subject of growing new interest. It is one of the most fundamental consequences of the interaction of light with matter and, increasingly, a very accessible common ground in which a broad scope of experiments that probe state-to-state unimolecular relaxation and fragmentation dynamics meet powerful new scattering theory methodologies. Now, a surge of new experimental activity is making use of a recently developed technique termed zero kinetic energy (ZEKE) photoelectron spectroscopy. Results presented at the Kreuth conference point to the emergence of several important new applications built on the unprecedented resolution of this method, including routine means for the accurate determination of thermochemically important ionization potentials, the convenient characterization of ion rovibrational level structure, including that of elemental clusters and weakly-bound molecular complexes, as well as the acquisition of fundamental data on state-detailed photoionization cross-sections and the dynamics of intramolecular relaxation.

ZEKE and conventional photoelectron spectroscopy

While unprecedented in resolution, typical ZEKE experiments share many of the objectives of conventional photoelectron spectroscopy. These measurements are based on Einstein's energy conservation expression, $KE = h\nu - \Delta E$, describing the photoelectric effect. Here, $h\nu$ is the photon energy of the monochromatic light, ΔE is the energy difference between rovibronic states of the neutral molecule and cation and KE is the kinetic energy of the ejected electron. Analysis of the photoelectron kinetic energy spectrum (KE) conveys information on the distribution of states in the cation (ΔE) and the dynamics of the ionization event itself. In conventional photoelec-

tron spectroscopy, however, one measures electrons of all kinetic energies, so that the resulting information on ion level structure is limited by the resolving power of the electron energy analyser, which is usually no better than 10 meV. An early refinement, termed threshold photoelectron spectroscopy (TPES), improves on this by excluding all but low-energy, near-threshold electrons ($KE \sim 0$ eV), and the spectrum of the cation (defined by ΔE) is then obtained by scanning the frequency of the light source ($h\nu$)¹. This method effectively matches the resolution limit of near-threshold electron-velocity discrimination with the few-meV photon-energy resolution typically available from vacuum ultraviolet (VUV) monochrometers. The ZEKE method, developed by Müller-Dethlefs and Schlag at the Technical University of Munich², introduces a key improvement that has opened the spectrum of threshold photoionization to interrogation at the resolution of modern pulsed lasers.

ZEKE spectroscopy

ZEKE spectroscopy exploits the time and frequency-resolution characteristics of laser light sources, together with the well known Stark shift of an ionization limit in an external electric field to register precisely threshold photoionization. It works because highly excited Rydberg levels very near ionization thresholds, which are bound in a field-free environment, yield free electrons when one applies an electrostatic field which induces a Stark shift larger than their binding energy. Typical ZEKE experiments achieve a resolution of 1 cm^{-1} (0.1 meV) or better by using small electric fields of 0.1–1.0 V cm^{-1} so that only Rydberg levels with principal quantum numbers $n > 150$ are field ionized. With pulsed laser photoexcitation, electrons arising from delayed field ionization can be distinguished in time from electrons created by prompt ionization processes, including rapid electronic and vibrational autoionization. By recording the electron signal due solely to delayed field ionization, one registers the spectrum exclusively of transitions to ionization thresholds, which precisely maps the rovibronic structure of the cation. Matching spectroscopic discrimination is achieved by using narrow-

bandwidth lasers in resonant multiphoton ionization (REMPI) schemes or, following frequency conversion to the VUV, in direct one-photon photoionization. Sequence congestion in rotationally resolved ZEKE is often eliminated by means of double-resonant or multiresonant optical selection. A typical ZEKE apparatus is shown in Fig. 1 (refs 2–4).

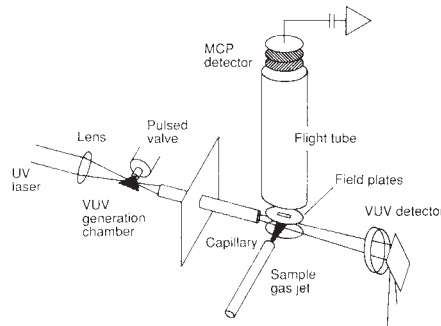


FIG. 1 Schematic diagram of a time-of-flight electron spectrometer used for studying the ionization dynamics and spectroscopy of cations produced by single-photon ionization using vacuum ultraviolet (VUV) light. The VUV radiation is produced from UV-laser light by nonlinear processes in free-jet expansions of various gasses and transported to the sample gas being studied by a long capillary light guide. Interaction of the VUV radiation and beam of sample molecules results in field-ionized electrons when an electric field pulse is applied across the two cylindrical field plates. The electrons are accelerated into the long field-free flight tube and analysed by their arrival time-of-flight at the detector (adapted from ref. 10).

Applications

The high resolution of the ZEKE technique provides a new means to explore longstanding questions concerned with the rotational state distributions of cations produced by photoionization. Indeed, before the development of laser-based photoelectron methods, only H_2 had been rotationally resolved by photoelectron spectroscopy. Such information offers a sensitive probe of ionization dynamics. Recent rotationally resolved ZEKE measurements for more than a dozen diatomic and small polyatomic molecules have highlighted the powerful way in which the spectra of cation internal states produced by threshold photoionization provide direct information on partial wave distributions of photo-

electrons. This experimental work has been paralleled by the rapid development of theoretical treatments using multichannel quantum defect theory (MQDT)⁵ and Hartree-Fock methods which explicitly include the rotational motion of the molecule⁶⁻⁸.

Several groups from Europe and the United States presented new results at the Kreuth meeting which suggest that final state interactions among Rydberg states near ionization thresholds can lead to anomalous rotational linestrengths in ZEKE spectra. Unusually high intensities for photoionizing transitions accompanied by large decreases in molecular core rotation have been found by the groups of Gerber (University of Freiburg, Germany), Müller-Dethlefs (Technical University of Munich, Germany), Softley (Oxford University, UK), Grant (Purdue University, USA), White (Brookhaven National Laboratory, USA) and Weisshaar (University of Wisconsin, USA) for diatomics, NO, O₂, HCl, OH, H₂, Na₂, N₂ and VO, and triatomics N₂O and NO₂. Figure 2 gives an example ZEKE spectrum illustrating this phenomenon for NO₂. In each case, spectral evidence has been seen to support Rydberg-Rydberg final-state interactions as the cause of these anomalous ZEKE intensities. The body of evidence clearly indicates that such final-state interactions are ubiquitous, and this development represents an exciting new opportunity to study multichannel interactions with high resolution.

Another major area of application of ZEKE spectroscopy represented at the Kreuth meeting is high-resolution vibrational spectroscopy of large organic cations. Conventional infrared spectroscopy of ions is limited by light source and detector considerations to narrow frequency ranges, whereas ZEKE techniques can provide a view at high resolution (≤ 1 cm⁻¹) over very wide intervals of ion vibrational energy. Novel applications reported in Kreuth include the work of Johnson and coworkers (SUNY-Stony Brook, USA) who have used ZEKE and (1+1') REMPI to probe the singlet-triplet state mixing responsible for intersystem crossing in pyrazine and pyrimidine, and the work of Kimura and coworkers (Institute for Molecular Science, Japan) who report isolating the cation vibrational spectrum of structural isomers in n-propylbenzene (*trans* and *gauche*) and chlorophenol (*cis* and *trans*). Also presented was the spectrum of the hydrogen bond in the isolated cationic phenol-water complex, which was obtained in a collaborative effort by Colson (Pacific Northwest Laboratory, USA) and Müller-Dethlefs and Schlag (Technical University of Munich, Germany). The ability of ZEKE spectroscopy to probe reactive transients was demon-

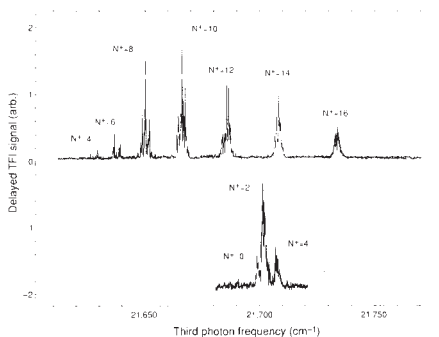


FIG. 2 ZEKE scans showing rotationally resolved threshold photoionization of NO₂ to produce NO₂⁺ in its ground vibrational state. Lower spectrum: photoionizing transitions to ion rotational states, $N^+=0, 2$ and 4 , from the doubly resonant selected $N^+=1$ rotational level of the $3p\sigma^2\Sigma_u^+$ Rydberg state. Upper spectrum: photoionizing transitions to allowed $N^+=10, 12, 14$, and 16 , plus anomalous $N^+=4, 6$ and 8 from $N^+=13$ in $3p\sigma^2\Sigma_u^+$. The fine structure for lower N^+ can be assigned to allowed transitions to Rydberg states of higher N^+ and lower principal quantum number, n , which in the neighbourhood of forbidden thresholds share intensity with (or equivalently, borrow lifetime from) high- n ZEKE Rydberg states (from ref. 11).

strated by the Weisshaar group who reported the cation vibrational spectrum of the benzyl radical, which is an important intermediate in many organic reactions, and the White group who studied the ionization dynamics of the OH radical, a central intermediate in combustion chemistry. Similar structural information can be obtained for even more exotic intermediates by extending high-resolution ZEKE methods to improve threshold discrimination in the photodetachment spectroscopy of anion clusters as well as anion precursors of atom plus diatom transition states.

In general, the well-defined properties of cations serve as a useful template upon which to disperse the dynamics of intramolecular relaxation in intermediate electronic states of neutral molecules. Time-resolved pump-probe photoelectron spectroscopic measurements reported by Knee (Wesleyan University, USA) and Reilly (University of Indiana, USA) illustrate how the time evolution of intensities in the cation vibrational spectrum, detected by photoionization, can be used to follow complex intramolecular relaxation processes in the first excited singlet states of large aromatic molecules. Time-dependent ZEKE spectra highlight the bath of vibrational modes that are responsible for intersystem crossing and vibrational relaxation with ultimate sensitivity and time-resolution far exceeding dispersed fluorescence methods.

New directions

Activity is growing in the fundamental application of ZEKE methods, and we can expect deeper inquiry into estab-

lished questions of rotational linestrengths and Franck-Condon factors for photoionization. The generality of ZEKE threshold photoionization spectroscopy as a method for ion vibronic structure determination is appealing and will see expanded use, especially for polyatomic systems and weakly-bound complexes in which the internal motions are characterized by low-frequency vibrations. One avenue as yet unexplored in this regard is the investigation of highly excited vibrational states of ions in parallel with, but potentially much broader than, similar such studies of high-lying vibrational structure and intramolecular coupling dynamics in neutrals conducted, for example, by stimulated emission pumping.

Methods for mass selection in cation-detected threshold photoionization point to new approaches for the study of internal state-selected ion-molecule reaction kinetics and dynamics.⁹ Techniques for the threshold identification and velocity separation of specific ion masses in specific internal states will be especially useful in cases where photoselection yields a mixture of ion internal states and/or masses, such as in sources for metal clusters and molecular complexes.

Finally, the discriminating power of ZEKE methods will allow spectroscopists to sort through regions of higher densities of states. We can expect to learn more, at higher resolution, about electronically excited states of cations and, with the extension of ZEKE detection to synchrotron light sources, foreseeable experiments will explore the rovibrational structure of doubly-charged molecules via the threshold photoionization of cations. □

Edward R. Grant, Department of Chemistry, Purdue University, West Lafayette, Indiana 47907, USA and Michael G. White, Chemistry Department, Brookhaven National Laboratory, Upton, New York 11973, USA. For more information, fill in reader service number 100.

1. Baer, T., Peatman, W. B. & Schlag, E. W. *Chem. Phys. Lett.* **4**, 243-247 (1969).
2. Müller-Dethlefs, K. & Schlag, E. W. *Ann. Rev. Phys. Chem.* **42**, 109-136 (1991).
3. Tonkyn, R. G., Winniczek, J. W. & White, M. G. *Chem. Phys. Lett.* **164**, 137-142 (1989).
4. Haber, K. S., Bryant, G. P., Jiang, Y., Grant, E. R. & Lefebvre-Brion, H. *Phys. Rev. A* **44**, 5331 (1991).
5. Child, M. S. & Jungen, Ch. *J. chem. Phys.* **93**, 7756-7766 (1990).
6. Lefebvre-Brion, H. *Chem. Phys. Lett.* **171**, 377-380 (1990).
7. Braunstein, M., McKoy, V., Dixit, S. N., Tonkyn, R. G. & White, M. G. *J. chem. Phys.* **93**, 5345-5346 (1990).
8. de Beer, E., de Lange, C. A., Stephens, J. A., Wang, K. & McKoy, V. *J. chem. Phys.* **95**, 714-715 (1991).
9. Zhu, L. & Johnson, P. M. *J. chem. Phys.* **94**, 5769-5771 (1991).
10. Tonkyn, R. G. & White, M. G. *Rev. Sci. Instrum.* **60**, 1245-1251 (1989).
11. Bryant, G. P., Jiang, Y., Martin, M. & Grant, E. R. (in preparation).

Tools of the trade

In the news this week — autorads with a white background, a silicon-based biosensor for neuroscience research and "clip art" designed specifically for biologists and chemists to improve the look of scientific presentation material.

TWO new Hyperpaper paper-based products have rolled off the Amersham production line. Both produce **autoradiographs with a white background** and are designed to offer a low-cost alternative to conventional autoradiograph film (*Reader Service No. 101*). Hyperpaper can be processed in the same manner as standard autoradiography film and requires no change to existing protocols. It can also be faxed or photocopied. Hyperpaper-³⁵S Sequencing is most sensitive to low-energy beta-emitting isotopes such as ³⁵S and ¹⁴C, and performs similarly to standard X-ray films in these applications, the company says. For high-energy beta and gamma emitters, Amersham recommends the use of Hyperpaper in those applications where resolution is more important than sensitivity. Hyperpaper-ECL Western is intended for enhanced chemiluminescence protein detection.

The Swedish company, Eka Nobel, has introduced a new line of Kromasil high-pressure, slurry-packed **chromatography columns**, which are available in sizes that range from 4.6-mm ID to 50.8-mm ID (*Reader Service No. 102*). The new columns are packed with Kromasil spherical silica (100 Å pore size) and in particle sizes ranging from 5 µm to 16 µm. Kromasil is a high-purity silica with a unique surface chemistry, producing high protein recovery and symmetrical peaks for basic compounds, the company says. It is mechanically stable up to 10,000 p.s.i. and is chemically stable in the pH range of 1.5 to 9.5.

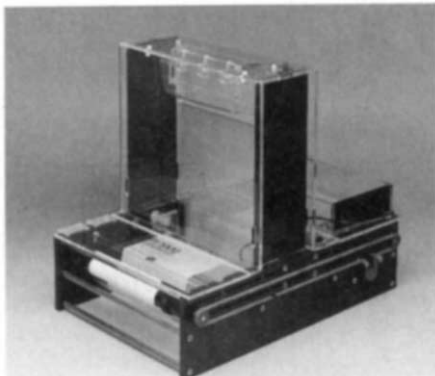
With its dual function, WARMCOLD from EquiBio can be used both as a **refrigerator and heater** (*Reader Service No. 103*). The device features a temperature range of -20 °C to +70 °C and, depending on your whim, it can be used for applications as diverse as hybridization procedures and cold storage. WARMCOLD uses Peltier technology and is microprocessor controlled. The storage area is completely sealed and so experiments can be performed either wet or dry.

Dealing with DNA

No taping is necessary when casting 55×70-mm gels in the gel-running tray of the new electrophoresis unit from Acrylglassverarbeitung (*Reader Service No. 104*). The unit is designed for the rapid and

precise **separation of PCR fragments** in 30–45 minutes. Just 15 ml of agarose solution is sufficient for a 14-sample gel, the company says. As the gel tray is transparent to UV, this enables photo-documentation of the gel directly within the unit. The electrophoresis unit, which is made of clear acrylic plastic, is supplied with two 7-well combs, end gates, platinum electrodes and safety power cords.

The new TE 2000 TwoStep **direct blotter** from Hoefer combines two functions in one instrument: sequencing gels are run and then blotted directly onto a nylon membrane that passes beneath the bottom edge of the gel (*Reader Service No. 105*). DNA sequences that are bound



TwoStep — two functions in one.

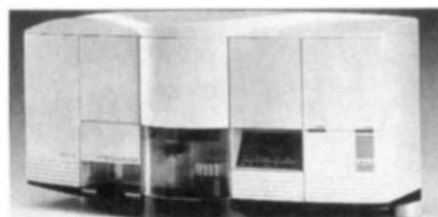
to the immobilizing membrane can be probed numerous times and developed with chemiluminescent or colourimetric detection systems, thus increasing the amount of information that can be obtained from a single gel, Hoefer says. The TE 2000 can also be used for the resolution of radiolabelled fragments. The TwoStep blotter includes a user-programmable microprocessor, which controls the membrane speed.

Applied Biosystem's CATALYST 800 **laboratory station for molecular biologists** automates Taq-cycle sequencing with fluorescent primers or terminators and is designed to increase productivity and reduce labour and reagent costs (*Reader Service No. 106*). After the samples and reaction mixes are manually loaded onto the CATALYST worksurface, users select a preprogrammed chemistry using a Macintosh computer. The laboratory station automatically performs the reaction set-up, thermal cycling and, in the case of dye primer chemistry, pooling and preci-

pitiation of the reaction products. These products are then manually loaded onto ABI's DNA sequencer for automated detection and analysis. According to ABI, 72 templates can be prepared in a day. "Intelligent" robotics and syringe-based, positive liquid displacement ensure that CATALYST is accurate even when pipetting sub-microlitre volumes, ABI says.

Made to measure

In June, the French company, Europhor Instruments, introduced its new **laser-induced fluorescence capillary electrophoresis system** — the IRIS 2000 (*Reader Service No. 107*). The IRIS 2000 has a concentration sensitivity from 10⁻¹¹ M to



Laser capillary electrophoresis system.

10⁻¹⁴ M, depending on the compounds analysed, the company says, and is particularly useful for the analysis of peptides and proteins. Design features of this fully automated system include: a temperature regulation system from 15 °C to 60 °C in steps of 0.1 °C, a sampler of 24 positions, a cassette to ease the manipulation of capillaries from 15 µm to 100 µm, two kinds of injection and the option of two lasers — Ar-ion and He-Cd — for added flexibility.

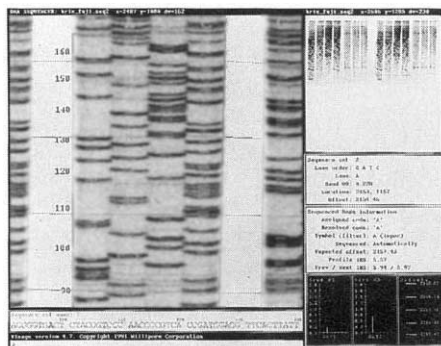
GlycoStation from Dionex brings the **automated analysis of glycoproteins** within the grasp of life scientists (*Reader Service No. 108*). The instrument is intended for a wide range of applications, including fundamental research, process monitoring, production quality control and routine analysis. It offers full automation of high-performance anion exchange separation and pulsed amperometric detection (HPAE-PAD) technology; complete system control using GlycoStation software; an interactive, self-teaching users guide and software that instructs the operator on how to take full advantage of HPAE-PAD methods; and, specially prepared monosaccharide, oligosaccharide and glycoprotein standards for system characterization and calibration.

The new Cytosensor Microphysiometer from Molecular Devices measures subtle changes in cell metabolism — both transient and sustained — and allows researchers to **track the response of cells to chemical and biological stimuli** within seconds or minutes (*Reader Service No. 109*). At the heart of the Cytosensor is a small chamber in which living cells are maintained near the light-addressable potentiometric sensor, or LAP sensor, surface. Measurements of cellular responses can be made non-invasively, with no harm to the cells, the company says. The same cells can be monitored continuously, with repeated applications of the cell-affecting agents. Molecular Devices says that the Cytosensor Microphysiometer shows promise in the characterization of neurotransmitters and other ligands, and also offers potential for a variety of other neuroscience applications, including the elucidation of biochemical pathways, understanding the role of nerve growth factors, drug discovery and toxicology testing.

Circular dichroism detection, fast time-resolved spectra (80,000 per second), zero-interrupt sequential mixing and global multi-parametric analysis are design features of the new DX.17MV biosequential **stopped-flow spectrofluorimeter** from Applied Photophysics (*Reader Service No. 110*). A ball mixer, low mass drive ram and high sample flow rates are used for protein folding/unfolding experiments where dense solutions like 6–8 M urea or guanidinium chloride are mixed with aqueous buffers. In addition, the sequential mixing capability can be used to study the initial phase of fast folding proteins — the second mix being used to refold before the slow second phase becomes significant, the company says. The DX.17MV was designed with researchers in biochemistry and molecular biology in mind.

Program profile

Millipore has developed a new **software program** called BioImage, which accepts and analyses images acquired by the Fuji BAS2000 Phosphor Imaging System (*Reader Service No. 111*). Fuji's system detects radiolabelled materials through an erasable phosphor imaging plate. The plate is designed to provide ultra-high sensitivity to radioactivity, linearity of response and high resolution. The image on the plate is read by a laser and converted into a computer file for analysis. BioImage software enhances the analysis capabilities of the BAS2000, Millipore says. Images from DNA sequencing, two-dimensional electrophoresis or restriction fragment length polymorphism applications can be accepted and analysed. BioImage automatically detects all bands or spots on an image,

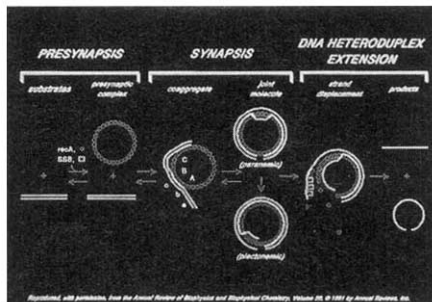


Biolmage analysis software.

quantitates the bands or spots in terms of molecular weight and optical density, and compares the results between separate images to point up the differences in experimental results.

ELISA READ is a new **computer interface program** from Meddata that is designed to allow any PC or compatible computer to control the data-capture functions of most makes of microplate reader (*Reader Service No. 112*). Once data are acquired, the disks can be analysed by two other Meddata programs, ELISA LITE or ELISA+. Options are included to save data in the *.PRN format used by Lotus-type spreadsheet programs, a *.TXT format for ASCII text transfer to database programs, and for computers running Microsoft Windows, in a format for the Excel spreadsheet. The organization of data files in ELISA READ has been designed for flexibility. For example, Meddata says that any number of plates can be recorded in a file. Once created, the file can be viewed and edited. To facilitate use, each file has an information box that can be accessed from the list of file names. The box contains information such as the date, time, file size, number of plates and number of readings per plate. As data accumulate, the pop-up information box should help researchers to find a particular assay run or select a particular plate for further analysis. Manual, automatic, "stacker" and fluorescence microplate readers are supported by this program.

Advanced Graphics Software is offering five Science Figure Packs containing **"clip art"**, a collection of scientific



Scientific 'clip art'.

symbols, drawings, maps and markers designed to improve the quality of presentations, reports, slides or publications (*Reader Service No. 113*). Each Figure Pack contains images that are relevant to a specific scientific area — biological cells, drugs, human anatomy, maps (US, Europe and global), and a general scientific figure pack with chemical and laboratory symbols. All five are designed for use with the company's SlideWrite Version 4 scientific presentation software for IBM PCs or compatible computers. All SlideWrite images are in vector format and so can be enlarged, reduced, coloured, tilted or rotated with no loss of resolution of the image, the company says. SlideWrite Figure Packs cost \$99 (US) each and are available on 3.5- or 5.25-inch disks.

Shelf stockers

I.D. NA agarose is a new low electroendosmosis, high gel strength agarose from FMC BioProducts developed specifically for the **resolution of DNA in DNA typing tests** (*Reader Service No. 114*). Typical applications include DNA identity testing using probes for variable-number tandem repeats, hypervariable repeats or restriction fragment length polymorphisms, as in forensic science, paternity testing, tissue typing and animal breeding. I.D. NA agarose is performance tested under rigorous DNA identity testing conditions to ensure lot-to-lot consistency, sharp DNA banding patterns and high-efficiency DNA transfers, FMC says.

Permanently bent plungers on 5- and 10- μ l syringes could be a thing of the past now that Hamilton has developed the **FLEXIT syringe product line** (*Reader Service No. 115*). Hamilton claims that its **FLEXIT syringe plungers**, which are made of inert, titanium-based alloy, remain straight even after being bent past the point at which stainless steel plungers would be ruined.

Earlier this month, Unicam launched Reflex — the company's **supplies service** for routinely used laboratory consumables, electrochemistry products, general spares and accessories — to a worldwide market (*Reader Service No. 116*). The Reflex service is designed to complement the company's range of analytical instruments for chromatography and spectroscopy. Coinciding with the announcement, Unicam offers a new catalogue and copies are available upon request. □

These notes are compiled by Diane Gershon from information provided by the manufacturers. To obtain further details, use the reader service card bound inside the journal. Prices quoted are sometimes nominal, and apply only within the country indicated.